

Verkkokyselytutkimuksesta monimuuttujamenetelmiin

Lukukausimaksukokeilu ja opiskelija-aines

Eija-Leena Koponen

Helsingin yliopisto
Valtiotieteellinen tiedekunta
Tilastotiede
Pro gradu -tutkielma
Huhtikuu 2012



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto – Fakultet/Sektion – Faculty Valtiotieteellinen tiedekunta		Laitos – Institution – Department Sosiaalitieteiden laitos	
Tekijä – Författare – Author Koponen, Eija-Leena			
Työn nimi – Arbetets titel – Title Verkkokyselytutkimuksesta monimuuttujamenetelmiin – Lukukausimaksukokeilu ja opiskelija-aines			
Oppiaine – Läroämne – Subject Tilastotiede			
Työn laji – Arbetets art – Level Pro gradu -työ		Aika – Datum – Month and year Huhtikuu 2012	Sivumäärä – Sidoantal – Number of pages 69 + liitteet
Tiivistelmä – Referat – Abstract <p>Suomessa vuoden 2011 syksyllä aloitettu lukukausimaksukokeilu EU- ja ETA-maiden ulkopuolisille maisterivaiheen opiskelijoille on herättänyt laajaa keskustelua ja huolta mm. ylioppilaskunnissa. Tämän tutkielman alkuperäinen tavoite oli selvittää kyselytutkimuksen avulla miten lukukausimaksukokeilu vaikuttaa opiskelija-ainekseen, sillä on todisteita siitä että maksujen käyttöönotto vaikuttaa negatiivisesti alemmista sosioekonomisista luokista tulevien korkeakouluun hakeutumiseen. Yllättävän pieni maksavien opiskelijoiden määrä muutti tutkimuksen otteen pilotiksi, jonka tavoitteena oli testata niin lomaketta kuin tutkimuskysymystä laajempaa jatkotutkimusta varten. Lisäksi tavoitteena oli hankkia tietoa ulkomaalaisten opiskelijoiden taustoista, asenteista ja rahankäytöstä.</p> <p>Tutkimuksen aineisto on kerätty Aalto-yliopiston maisteriohjelmien ulkomaalaisilta opiskelijoilta tutkimusta varten laaditulla verkkokyselylomakkeella. Verkkokyselylomakkeella on monta muista tiedonkeruutavoista poikkeavaa ominaisuutta, joita käsitellään tutkielmassa. Lisäksi vastauskadon ehkäisyyn perehdytään kyselylomakkeen rakentamisvaiheessa. Kyselyn muuttujat mittaavat opiskelijoiden sosioekonomisen taustan lisäksi motivaatiota hakeutua Suomeen opiskelemaan ja asenteita lukukausimaksuja kohtaan. Aineistoa analysoidaan esimerkinomaisesti faktorianalyysillä sekä ryhmittelyanalyysillä. Faktorianalyysin teko aloitetaan konfirmatorisella lähestymistavalla jatkaen eksploratiivisella otteella. Ryhmittelyanalyysi muodostetaan kahta faktoripistemuuttujaa käyttäen hierarkisella ja ei-hierarkisella metodilla. Ryhmiä pyritään kuvailemaan taustamuuttujien avulla. Lisäksi menetelmien teoriaa esitellään tutkielmassa laajemmin.</p> <p>Tutkielman tulokset monimuuttujamenetelmien osalta jäivät vaatimattomiksi ensinnäkin tutkimuskysymysten muututtua pilottimaiseksi kohderyhmän laajentamisen myötä. Lisäksi aineiston muuttujien jakaumat olivat liian vinoja hyvien analyysien tekemiseksi. Näin ollen tulokset toimivat esimerkkeinä, jos tutkimusta päätetään uusiksi. Pilottitutkimuksena tutkimus on onnistunut ja ja sen avulla saatiin paljon tietoa kyselyn teosta sekä opiskelijoista. Tutkimusta uusittaessa on syytä käyttää lyhyempää kyselylomaketta ja ottaa paremmin huomioon kaikki erityisryhmät ulkomaalaisten tutkinto-opiskelijoiden joukossa. Kyselytutkimuksen vastausprosentti oli hyvä (n. 46 %), mikä johtunee opiskelijoita kiinnostavan aiheen lisäksi, siitä että tutkimukseen osallistumisen porkkanana oli 100 euron lahjakortin arvonta. Kysely suunniteltiin ja toteutettiin noudattaen kirjallisuudessa esitettyjä ohjeita ja kyselylomakkeeseen vastattiin tasaisesti, joten myös itse kysely oli vähintäänkin melko onnistunut. Valitettavasti keskeyttäneistä ei ollut tilaisuutta saada informaatiota ja vastauskatoanalyysi jäi tyngäksi.</p> <p>Tutkimuksen tuloksia voi sinällään käyttää hyväksi opiskelijoiden edunvalvonta työssä myös muun kuin lukukausimaksukokeilun parissa. Esimerkiksi ulkomaalaisten opiskelijoiden sosioekonomista taustaa tai rahankäyttöä ei ole ennen tätä tutkittu vastaavalla laajuudella. Lisäksi tämän pilottitutkimuksen ja sen tulosten perusteella Aalto-yliopiston ylioppilaskunnalla (AYY) on mahdollisuus toteuttaa seuraavina vuosina vastaavanlainen kysely omille opiskelijoilleen, tai laajentaa tutkimusta koko Suomea koskevaksi ja saavuttaa näin parempaa tietoutta koskien lukukausimaksukokeilua.</p>			
Avainsanat – Nyckelord – Keywords verkkokyselytutkimus faktorianalyysi ryhmittelyanalyysi lukukausimaksukokeilu			

Sisällys

1	Johdanto	5
2	Taustaa: korkeakoulutuksen rahoituksesta	8
2.1	Korkeakoulutus murroksessa	8
2.1.1	Korkeakoulutuksen kasvavat kulut	9
2.2	Kuka maksaa korkeakoulutuksen?	10
2.2.1	Koulutuksen taloudelliset vaikutukset	10
2.2.2	Periaatteellisia näkökulmia korkeakoulutuksen rahoitukseen	12
2.2.3	Korkeakoulutuksen rahoitusmalleja	15
2.3	Lukukausimaksukokeilu Suomessa	16
2.3.1	Lukukausimaksukokeiluun johtaneita tekijöitä	17
2.3.2	Kokeilun toteutus	18
2.3.3	Kokeilua kohtaan esitetty kritiikki	19
2.3.4	Lukukausimaksukokeilu Aalto-yliopistossa	20
3	Kyselytutkimuksen suunnittelu ja toteutus	23
3.1	Kyselylomakkeen suunnittelu	24
3.1.1	Mittausmalli	24
3.1.2	Kehikkoperusjoukko ja sen asettamat vaatimukset	27
3.2	Verkkolomake tiedonkeruuvälineenä	28
3.2.1	Huomioon otettavaa verkkolomaketta laatiessa	29
3.2.2	E-lomakepalvelu	30
3.3	Vastauskadon ennaltaehkäisy	31
3.3.1	Kyselytutkimukseen liittyvät yleiset tekijät	32
3.3.2	Kysymysten ominaisuudet	33
3.3.3	Kyselylomakkeen ja kysymysten testaus	35
3.4	Aineisto	36
3.4.1	Aineiston keruu	36
3.4.2	Saadun aineiston kuvaus	36
3.4.3	Puuttuneisuus	39
3.4.4	Aineiston muokkaus	40
4	Monimuuttujamenetelmät	42
4.1	Faktorianalyysi	42
4.1.1	Faktorianalyysimalli	42
4.1.2	Suurimman uskottavuuden faktorointi	43
4.1.3	Rotaatiomenetelmät	44
4.1.4	Faktoripisteet	46
4.2	Ryhmittelyanalyysi	47

4.2.1	Etäisyys- ja erilaisuusmitat	48
4.2.2	Hierarkinen ryhmittely	50
4.2.3	Ei-hierarkinen ryhmittely	51
5	Kyselyn tulosten analysointi monimuuttujamenetelmillä	54
5.1	Faktorianalyysin tulokset	54
5.1.1	Faktoripisteet	56
5.2	Ryhmittelyanalyysi faktoripisteillä	58
5.2.1	Hierarkinen ryhmittely	58
5.2.2	Ei-hierarkinen ryhmittely ja ryhmien tulkinta	60
6	Johtopäätökset	64
	Lähteet	66
	Liite A Alustava tutkimussuunnitelma	70
	Liite B Sitoumus	72
	Liite C Saatekirje	73
	Liite D Kyselylomake	74
	Liite E Perustunnuslukuja	83
	Liite F Taustamuuttujia ryhmittäin	95
	Liite G Ohjelmakoodit	97

1 Johdanto

Korkeakoulutus ja sen maksullisuus on herättänyt vilkasta keskustelua viime vuosina niin mediassa kuin päättäjien ja opiskelijoiden keskuudessa. Suomessa aloitettiin korkeakouluille vapaaehtoinen lukukausimaksukokeilu englanninkielisissä maisteriohjelmissa EU- ja ETA-maiden ulkopuolisille opiskelijoille syksyllä 2011, ja se jatkuu vuoteen 2014 saakka. Maksuttomuuteen tottuneet ja etunsa saavuttaneet opiskelijat ovat kauhistuneita ajatuksesta, että päättäjät ajavat opintojen maksullisuutta ja maksukokeilun oletetaan olevan vain askel kohti kaikkia koskevia lukukausimaksuja. Lisäksi pelätään, että suomalaisten kansainvälistyminen pysähtyy ja taantuu, jos lukukausimaksut kohdistetaan ulkomaalaisille opiskelijoille. Päättäjät ovat puolestaan huolissaan vähenevästä työvoimasta ja verotuloista suurten ikäluokkien jäädessä eläkkeelle. Tulevaisuuden veronmaksajat eli opiskelijat pitäisi saada nopeammin valmistumaan sekä työelämään. Lisäksi korkeakoulutuksen rahoitusta tulisi muuttaa yhä enemmän yksityiseen rahoitukseen pohjautuvaksi, jotta valtion vastuu rahoituksesta kevenisi. Myös huomio siitä, että lisääntyvä kansainvälistymisen myötä ulkomaalaisten opiskelijoiden on helpompi käyttää hyväksi suomalaista järjestelmää, eli hankkia ilmainen koulutus ja palata kotimaahansa tietotaidon ja tulevien verotulojen kanssa, on lisännyt paineita koulutuksen maksullisuus keskustelulle.

Tämä pro gradu -tutkielma tutkii lukukausimaksukokeilun vaikutuksia kyselytutkimuksen avulla. Työssä esitellään kyselytutkimuksen tekoa internetissä, aineiston keruuta ja sekä aineiston analysointia monimuuttujamenetelmillä. Internetissä tehtävä verkkokyselytutkimus tarjoaa paljon uusia mahdollisuuksia kyselylomakkeiden maailmassa. Verkkolomakkeet eroavat hyvin paljon perinteisistä postitettavista paperilomakkeista sekä puhelinhaastatteluista mm. interaktiivisen luonteensa, täysin uudenlaisen ympäristönsä ja itseohjautuvuutensa vuoksi (Couper, 2008, s.28). Tekemässäni kyselytutkimuksessa on monia muuttujia eli kysymyksiä samasta aiheesta, sillä subjektiivisiin kysymyksiin liittyy enemmän epävarmuutta, jolloin on hyvä tehdä aiheesta useampi kysymys (Laaksonen, 2009). Monimuuttujamenetelmillä aineistoa tiivistetään selkeämmän tulkinnan löytämiseksi. Faktorianalyysin avulla ryhmitellään muuttujia samankaltaisiin ryhmiin kun taas ryhmittelyanalyysillä ryhmitellään alkioita eli vastaajia mahdollisimman homogeenisiin ryhmiin. Ryhmien avulla voidaan kuvata mm. erilaisia opiskelijatyyppejä.

Tutkimuksen aihe syntyi Aalto-yliopiston ylioppilaskunnan (AYY) toimeksiannon pohjalta liittyen mainittuun lukukausimaksukokeiluun. AYY etsi graduntekijää selvittämään, miten maksukokeilu vaikuttaa Aalto-yliopistoon valikoituvaan opiskelija-ainekseen sekä opiskelijoiden sosioekonomisiin taustoihin. Tarkoituksena oli verrata 2010 ja 2011 aloittaneita opiskelijoita toisiinsa taustan, taloudellisen tilanteen ja asenteiden suhteen. AYY toimi tässä tutkimuksessa siis aiheenrajaajana, materiaalin tarjoajana sekä apuna mm. lomakkeen suunnittelussa.

AYY:n huolena on, että maksukokeilun myötä kouluihin syntyy kahden kerroksen

väkeä. Maksukokeilussa mukana oleville maisteriohjelmille on mm. budjetoitu enemmän rahaa opetuksen kehittämiseksi kansainvälisen kilpailun kestäväksi. Maksukokeiluun osallistuvien ohjelmien opiskelijoille tarjotaan siis parempaa opetusta ja ohjausta sekä parempia palveluja kun muissa maksuttomissa ohjelmissa opiskelijoita kohdellaan toisen luokan ilmaisopiskelijoina. Lisäksi pelkona on, että opiskelijajoukko homogenisoituu liiaksi sisältäen vain niitä ulkomaalaisia opiskelijoita, joilla on varaa maksaa. (Henkilökohtainen viestintä, 10.3.2011.)

Kuten luvussa 2 kirjoitan, on olemassa empiiristä näyttöä siitä, että lukukausimaksujen käyttöönotto ja/tai lukukausimaksujen korottaminen vaikuttavat negatiivisesti alemmista sosioekonomisista luokista tulevien koulutukseen osallistumiseen. AYY pelkää, että maksukokeiluun kuuluvasta stipendijärjestelmästä huolimatta, huonompiosaisilla ei ole enää varaa tulla jo elinkustannuksiltaan kalliiseen Suomeen. Myös päättäjien mielestä ulkomaalaiset ovat tärkeä osa Suomen ja suomalaisten kansainvälistymisprosessia. Ulkomaalaisten vähetessä menetetään mahdollisuus Suomen tulevaisuuden kannalta tärkeään ulkomaalaiseen osaavaan työvoimaan sekä tuleviin veronmaksajiin. Toisaalta ulkomaalaiset opiskelijat tulevat mm. International Student Barometer (ISB)-kyselyn mukaan Suomeen pääosin ilmaisen koulutuksen perässä ja jäävät harvemmin Suomeen töihin (Aallon kv-opiskelijoiden tyytyväisyystutkimus, henkilökohtainen viestintä, 10.3.2011.). Tällöin he vievät mukanaan arvokkaan koulutuksen, jonka suomalaiset veronmaksajat ovat maksaneet.

Opetus- ja kulttuuriministeriön asettama työryhmä arvioi lukukausimaksukokeilua ja sen onnistumista, mutta ei aio selvittää lukukausimaksuista koituvia vaikutuksia opiskelijainekseen. AYY pitääkin asian selvittämistä erittäin tärkeänä opiskelijoiden edunvalvonnan kannalta. Tutkielman tavoitteena on siis tuottaa objektiivista tietoa lukukausimaksukokeilun vaikutuksista Aalto-yliopistossa. (Henkilökohtainen viestintä, 10.3.2011)

Niin kuin oli odotettavissa Tanskan ja Ruotsin vastaavien kokemusten perusteella, hakijamäärät laskivat lukukausimaksujen käyttöönoton jälkeen Aalto-yliopistossa. Hakijamäärien lasku jäi kuitenkin vain 12 %:iin verrattuna vuoteen 2010 (Aalto-yliopiston hakijapalvelut, 2011). Vuoden 2011 valinnoissa Aalto-yliopiston maksukokeiluun osallistuviin maisteriohjelmiin hyväksyttiin 81 ulkoeurooppalaista opiskelijaa. Heistä 40 vastaanotti opiskelupaikan määräaikaan mennessä ja ilmoittautui läsnä olevaksi. Varsinkin stipendittä jääneet peruivat opiskelupaikkojaan. Lopulta vain kolme opiskelijaa maksaa täydet lukukausimaksut ja yhdeksän opiskelijaa maksaa puolet maksuista (taulukko 2.4).

Näin pieni maksavien opiskelijoiden määrä vaikutti myös tutkimuksen toteutukseen, sillä vertailtavien ryhmien koko ei olisi riittänyt merkitsevien tilastollisten analyysien tekoon, vaikka kaikki kohderyhmään kuuluvat olisi saatu osallistumaan kyselyyn. Opiskelijamäärien varmistuttua tutkimusta päätettiin muokata pilottitutkimukseksi. Tämä tarkoitti sitä, että tutkimuksen ideaksi muodostui tutkimusasetelman sekä kyselylomakkeen testaus. Laadittu kyselylomake lähetettiin laajemmalle kohderyhmällä kuin aluksi suunniteltiin, joten saadut tulokset eivät vastaa alkuperäistä kysymyksen asettelua ja niihin tulee suhtautua suuntaa-antavina. Työn tarkoituksena on siis selvittää ulkomaalaisten opiskelijoiden suhtautumista maksuihin, heidän sosioekonomista taustaansa kyselylomakkeen avulla, analysoida tuloksia monimuuttujamenetelmillä sekä tarjota välineitä laajemman tutkimuksen tekoon tulevaisuudessa.

Tutkielmassa avataan ensin korkeakoulutuksen rahoitukseen liittyviä kysymyksiä mm. taloudellisten vaikutusten muodossa ja syvennyttään Suomessa ja etenkin Aalto-yliopistossa käynnissä olevaan lukukausimaksukokeiluun. Luvussa 3 perehdytään kyselytutkimuk-

sen metodiikkaan painottaen internetissä tehtävää verkkokyselyä ja käytetään esimerkkinä työtä varten tehtyä kyselylomaketta. Lisäksi luvussa kuvaillaan lomakkeen avulla kerättyä aineistoa. Valitsemieni monimuuttujamenetelmien, faktorianalyysin ja ryhmitelyanalyysin, teoriaa esitellään luvussa 4 ja aineistosta analysoituja tuloksia esitellään luvussa 5. Luku 6 esittelee tutkimuksen johtopäätöksiä ja arvioi pilottitutkimuksen onnistumista.

2 Taustaa: korkeakoulutuksen rahoituksesta

Tässä luvussa esittelen korkeakoulutuksen rahoitusta ja rahoitusmallien kehittymistä lähivuosikymmeninä, päätyen Suomessa käynnissä olevaan lukukausimaksukokeiluun. Esittelen rahoituspohjan laajentamisen eri näkökantoja koulutuksen taloudellisten vaikutusten näkökulmasta ja tämän jälkeen avaen lyhyesti maksuttomuuden ja maksullisuuden puolustajien periaatteellisia näkökulmia. Käytän Suomen tämänhetkistä tilannetta esimerkkinä ja pyrin vertaamaan tilannetta muissa maissa vallitseviin käytäntöihin.

2.1 Korkeakoulutus murroksessa

Korkeakoulujärjestelmä on läpikäynyt mittavia muutoksia viimeisen 50 vuoden aikana mm. koulutuksen kysynnän kasvaessa ja teollisten maiden siirtyessä yhä enemmän tietopohjaiseen osaamiseen (mm. Asplund, 2006; Johnstone, 2004). Koulutuksen kysynnän räjähtävä kasvu on tuonut haasteita kasvavine menoineen, mutta toisaalta korkean koulutusasteen myötä Suomesta on kasvanut menestyvä ja innovatiivinen, edistyneen teknologiateollisuuden maa (Puustelli, Niskanen, Sund & Rajala, 2009, s.10).

Kehittyneissä maissa korkeakoulututkinnon hankkineiden osuus väestöstä on kasvanut jyrkästi 1950-luvulta lähtien. Suomessa opiskeli vuonna 1955 yliopistoissa tai vastaavissa laitoksissa 16 628 opiskelijaa (Eicher, 1998, s.31). Opetusministeriön ylläpitämän KOTA-tietokannan mukaan oli opiskelijoita noin viisikymmentä vuotta myöhemmin vuonna 2009 jo 165 000, eli melkein kymmenkertainen määrä (taulukko 2.1). OECD:n (2009) mukaan korkea-asteen tutkinnon suorittaneiden osuus 25–34-vuotiaista suomalaisista on n. 40 % ja vastaavasti 55–64-vuotiaista alle 30 %. Korkeakoulutuksen hankkineiden osuus on siis noussut selvästi niin Suomessa kuin muuallakin teollistuneissa maissa. Lisäksi ulkomaalaisten osuus Suomen korkeakouluopiskelijoista on

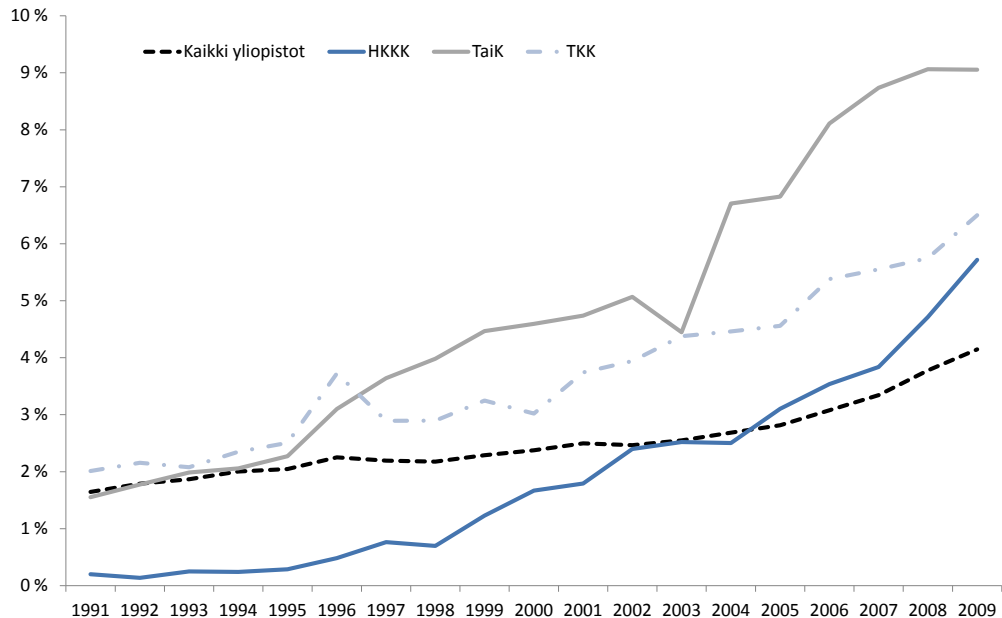
Taulukko 2.1. Opiskelijat

Vuosi	Opiskelijoita, lkm	
	Kaikki	Ulkomaalaiset
1991	115 368	1 899
2009	168 475	6 984
Muutoskerroin	1,46	3,68

Lähde: KOTA-tietokanta <https://kotaplus.csc.fi>

kasvanut tasaisesti ja se on yli kolminkertaistunut vuodesta 1991 vuoteen 2009. Nykyisen Aalto-yliopiston kouluissa (HKKK, TaiK ja TKK) lukumäärä on viisinkertaistunut. OECD:n (2011) mukaan Suomessa kolmannen asteen opiskelijoista 3,7 % on ulkomalaisia. Etenkin nykyinen Aalto-yliopiston Taiteiden ja suunnittelun korkeakoulu, entinen

Taideteollinen korkeakoulu TaiK, houkuttelee runsaasti ulkomaalaisia opiskelijoita ja ulkomaalaisten opiskelijoiden osuus on jopa 9,1 %. Kuvasta 2.1 nähdään, että ulkomaalaisten osuus on suurempi kaikissa nykyisen Aalto-yliopiston kouluissa kuin yhteensä kaikissa Suomen yliopistoissa.



Kuvio 2.1. Ulkomaalaisten osuus opiskelijoista. Aineisto: KOTA-tietokanta

Koulutuksen kysynnän kasvaessa ja koulutukseen käytettävien varojen paisuessa on alettu pohtimaan korkeakoulutuksen rahoitus pohjaa ja sen mahdollista laajentamista. Rahoitus pohjan laajentamisella tarkoitetaan valtion osuuden vähentämistä sekä yksityisen rahoituksen osuuden kasvattamista esimerkiksi ottamalla käyttöön lukukausimaksut. Yleisesti kerättävien verorahojen käyttöä puoltavat korkeakoulutuksen välittämät hyödyt yhteiskunnalle kuten innovaatiot ja kouluttautuneiden tulevaisuudessa maksamat korkeammat verotulot korkeampien palkkojen myötä. Rahoitus pohjan laajentamista opiskelijaa koskevaksi, eli lukukausimaksuja, puoltavat yksilön saavuttama korkeampi tulotaso, johon koko yhteiskunta antaa panoksensa verojen muodossa. Tämä koskee etenkin ulkomaalaisia opiskelijoita, jotka usein palaavat takaisin kotimaahansa.

2.1.1 Korkeakoulutuksen kasvavat kulut

Mikään koulutus ei ole ilmaista, vaikka Suomessa maksuttomasta koulutuksesta välillä puhutaankin, vaan varoja kuluu mm. opettajien palkkoihin, opiskelijapalveluihin ja tilojen käyttöön. Koulutuksen maksuttomuus opiskelijan näkökulmasta johtuu siitä, että useimmat teollisuusmaat tukevat, tai ovat tukeneet, korkeakoulutusta yleisesti kerättävistä verorahoista. Valtionavulla on perinteisesti haluttu taata kaikille kansalaisille yhtäläinen mahdollisuus osallistua koulutukseen omasta tai perheen tulotasosta riippumatta. Näin itse opiskelu on ollut pitkään opiskelijalle maksutonta tai lähes maksutonta. (Hölttä, 1998; Henttonen, 2008, s.27)

Koulutukseen käytettävät varat ovat paisuneet vuosien saatossa koulutuksen kysynnän kasvaessa. Vuonna 1960 Suomessa käytettiin korkeakoulutukseen 0,3 % bruttokan-

santuotteesta (BKT) ja vuonna 1980 jo 1,04 % (Eicher, 1998, s.33). Unescon (2009) mukaan vuonna 2000 osuus oli jo 2,0 % BKT:stä. Suomen valtion budjetissa tämä summa on tarkoittanut 2000-luvulla noin 4 %:n osuutta (Tilastokeskus). OECD:n (2011, Taulu B2.3) mukaan Suomi käyttää, yhdessä Norjan ja Tanskan kanssa, muihin OECD -maihiin verrattuna eniten julkisia varoja korkeakoulutuksen tukemiseen (1,6 % BKT:stä) ja viidenneksi eniten varoja, jos otetaan mukaan yksityinen rahoitus. Rahoituksen kasvu ei ole kuitenkaan pysynyt muiden OECD-maiden tasolla vuodesta 1995 vuoteen 2004 verrattuna, vaan on selvästi OECD:n keskiarvoa (55 %) matalampi (28 %) (OECD, 2007b, Taulu B1.5).

Kansantaloustieteilijät ovat jo pitkään puhuneet koulutuksen rahoituspohjan laajentamisen puolesta (mm. Barr, 2004). Edellä mainittu koulutuksen kysynnän räjähtävä kasvu yhdistettynä teollistuneiden yhteiskuntien ikärakennemuutokseen on luonut lisää paineita valtion korkeakoulutusta koskevan rahoitusvastuun keventämiseen (Barr, 2004). Taloudelliset alamäen jaksot, ensin 1990-luvun alussa ja nyt vuodesta 2008 lähtien, ovat lisänneet tarvetta leikata valtion kustannuksia siirtämällä vastuuta korkeakoulutuksen rahoituksesta yksityisille rahoittajille ja koulutuksesta nauttiville yksilöille (mm. Hölttä, 1998; Puustelli ym., 2009). Eläke- ja terveystaloudellisten nousussa ja säästöpainneiden kasvaessa, poliittisesti helpompi kohde leikkauksille on korkeakoulutus verrattuna esimerkiksi sairaanhoitoon tai eläkkeisiin. Lisäksi koulutuksen tuomat hyödyt ovat vaikeasti havaittavissa sekä mitattavissa (Asplund, 2006, s.38).

2.2 Kuka maksaa korkeakoulutuksen?

Korkeakoulutuksen rahoituspohjasta ja sen laajentamisesta on länsimaissa kiistelty jatkuvasti parin vuosikymmenen ajan. Korkeakoulutuksella on taloudellisia vaikutuksia niin yhteisölle kuin yksilöllekin. Lisäksi periaatteelliset näkökulmat jakavat kannanottajia eri leireihin.

2.2.1 Koulutuksen taloudelliset vaikutukset

Koulutuksen vaikutuksia voidaan vertailla makrotaloudellisesta eli yhteiskunnallisesta tai kokonaiskuvallisesta näkökulmasta sekä mikrotaloudellisesta näkökulmasta eli yksilön kannalta. Valtiota hyödyttävät makrotaloudelliset vaikutukset kuten korkeammat verotulot ja negatiiviset mikrotaloudelliset vaikutukset kuten menetetyt ansiotulot osoittavat valtion vastuun olevan paikallaan korkeakoulutuksen rahoitusjärjestelmässä. Toisaalta yksilön saama hyöty korkeakoulutuksesta on huomattava ja lisäksi valtio menettää verotuloja pitkän koulutuksen aikana, joten yksilön olisi näin loogista osallistua koulutuksen kuluihin.

Makrotaloudelliset vaikutukset

Korkeasti koulutetusta väestöstä on etua valtiolle mm. siksi, että korkeasti koulutetut ansaitsevat keskimäärin enemmän kuin alemmin koulutetut¹ (Ala-Panula, Bailey, Löppönen & Rautanen, 2009, s.45). Progressiivisen verotuksen maissa, kuten Suomessa,

¹Suomen virallinen tilasto (SVT): Palkkarakenne [verkkojulkaisu]. http://www.stat.fi/til/pra/2009/pra_2009_2010-11-12_tie_001_fi.html

korkeammat palkat johtavat korkeampaan veroprosenttiin, joten korkeasti koulutettu yksilö tuo enemmän rahaa valtion verokassaan (Henttonen, 2008, s.29). OECD:n (2011, s.138) mukaan korkeakoulutuksen hankkineet ansaitsivat OECD-maissa keskimäärin 50 % enemmän kuin toisen asteen koulutuksen saaneet. Suomessa eroa oli vain 20 %, joka johtuu vahvasta progressiivisesta verotuksesta ja tasaisesta palkkarakenteesta. Tasaisella palkkarakenteella tarkoitetaan sitä, ettei korkeakoulutuksen tuoma lisäansio ole tuntuva verrattuna toisen asteen koulutuksen hankkineeseen. Korkeakoulutettujen korkeampi palkkataso tuo myös tuloja valtion kiristuun kasvavan ostovoiman ja kulutuksen välityksellä, sillä jokaisesta tuotteesta ja palvelusta maksetaan myös veroa.

Asplundin (2006, s.9) mukaan koulutuksella on myös positiivinen vaikutus tuottavuuden ja talouden kehitykseen ja siten kaikkien kansalaisten ja yhteiskunnan hyvinvointiin. Vaikutusten voimakkuuksista on kuitenkin ristiriitaisia tutkimuksia, niin kuin myös syistä ja vaikutuskanavista. Keskimääräisen koulutustason nousu tuottaa koko yhteiskuntaa hyödyntäviä ulkoisvaikutuksia (Asplund, 2006, s.18). Ylemmillä koulutuksen asteilla (korkeakoulutus) hyöty tulee lähinnä teknologisen kehityksen ja tuottavuuden kasvun muodossa. Koulutuksen ajatellaan siis kiihdyttävän taloudellista kasvua esimerkiksi vaikuttaen työntekijän tuotannon tehoon ja sitä kautta kasvuun. Suomessa koulutustason nousu on ollut esimerkiksi tärkeä kasvuvauhtia ylläpitävä tekijä (emt.).

Muita koulutuksen epäsuoria vaikutuksia on hankala tunnistaa ja määrittää, mutta siihen on esitetty eri keinoja. Koulutuksen vaikutusta on mitattu esimerkiksi tuottoasteella, mikä kertoo koulutusinvestointien yhteiskunnalle antamasta taloudellisesta hyödystä (de la Fuente, 2003; Asplund, 2006, s.42–43). Yhteiskunnalle koitua hyöty muodostuu de la Fuenten (2003) mukaan kolmenlaisesta vaikutuksesta, joita ovat kokonaistuotanto, teknologinen kehitys sekä työllisyys. OECD (2010, Taulu A8.4.) laskee samankaltaista hyötyä rahassa. Koulutuksen yhteiskunnallista hyötyä laskettaessa menoihin kuuluvat koulutuksen aikaiset verotulon menetykset, sosiaaliset tuet ja koulutuksen julkiset menot. Tuloihin lasketaan korkeammista ansioista johtuvat verotulot ja säästöt sosiaalisten tukien määrissä. Suomessa miehen hankkima korkeakoulutuksen nettotuotto on \$ 107 500 (84 400 euroa) ja naisen taas \$ 41 700 (32 700 euroa). Miehen hankkiman koulutuksen hyöty ylittää Suomessa OECD:n keskiarvon kun taas naisen alittaa sen.

Tuottoasteen suuruus kertoo siis julkisten koulutusinvestointien tehokkuuden, joka Suomessa on 8,3 %. Suomi sijoittuu näin alhaisimmalle tuottotasolle Euroopan Unionissa. Toisaalta tuottoaste ei kerro kannattaako maan sijoittaa koulutukseen vaan siitä, mikä on kyseisessä maassa tuotto muiden sijoituskohteiden välillä. Asplund mainitsee esimerkkinä fyysisen pääoman, jonka tuottoaste on 2,5 – 4 prosenttiyksikköä pienempi kuin koulutuksen tuottoasteen. Tämä tuottoetu on taas Suomessa EU-maiden suurin, mikä tekee koulutuksesta yhteiskunnan kannalta houkuttelevan investoinnin. (Asplund, 2006, s.44.)

Henttonen (2008, s.17-18) perustelee alhaisten koulutusinvestointien tuottojen syyksi sen, ettei Suomessa korkeakoulutus vaikuta merkittävästi palkkatasoon. Toisaalta suomalaisilla ei usein ole isoja lainoja maksettavanaan heti koulutuksen jälkeen. De la Fuente (2003) esittääkin, että parempi mitta yksityisen ja yhteiskunnallisen tuoton väliseen erotukseen olisi vertailu yksityisen ja yhteiskunnallisen koulutuksen tuottaman palkkapreemion eli tutkintojen antamien tuottojen välillä. Näin kaikissa pohjoismaissa tukea tulisi jopa lisätä. Tulos selittynee myös Henttonen mainitsemalla matalalla palkkarakenteella.

Asplundin (2006, s.47) mukaan korkeaa koulutustasoa hyödyntäen on luotu uusia,

tuottavampia teknologioita ja etenkin koulutuksen ansiosta nämä on onnistuttu ottamaan käyttöön. Tätä kutsutaan kommunikointivaikutukseksi. Suomessa tapahtunut perinpohjainen rakennemuutos koulutuksessa on toteutunut tuottavuutta ja kilpailukykyä vahvistavalla tavalla. Koulutus on siis erilaisten vuorovaikutusmekanismien vuoksi tärkeä, muttei riittävä yksinään kasvun moottoriksi. (Asplund, 2006, s.49.)

Koulutuksella on siis suoria ja epäsuoria makrotaloudellisia vaikutuksia. Vaikutuksista, niiden mekanismeista sekä vahvuuksista kiistellään useilla tahoilla. Voidaan kuitenkin todeta, että korkeakoulutuksen antelias tukeminen on auttanut ainakin osaksi Suomea menestyksen tiellä ja korkeakouluttaminen on Suomessa kannattava sijoitus.

Mikrotaloudelliset vaikutukset

Korkeakoulutuksen hankkimisesta on yksilölle kiistatonta hyötyä. OECD:n (2007a) mukaan korkeakoulutuksen hankkineiden työttömyysprosentti oli Suomessa vuonna 2007 selvästi alhaisempi (4,4 %) kuin pelkän peruskoulutuksen hankkineiden (10,7 %). Lisäksi korkeakoulutettujen palkkataso sekä palkkatason kehitys ovat korkeampia kuin peruskoulun käyneillä (Ala-Panula ym., 2009, s.45). Toisaalta edellä mainitut tekijät vaihtelevat paljon ammattialoittain. Esimerkiksi sosiaali- ja terveysala on huomattavasti matalapalkkaisempaa alaa kuin lainopillinen ala. Selkeiden rahallisten tekijöiden lisäksi korkeakoulutus tuo hankkijalleen inhimillistä hyvinvointia ja pääomaa parempien taloudellisten olojen, aktiivisemmän poliittisen osallistumisen sekä paremman terveyden myötä (Ala-Panula ym., 2009).

Koulutuksesta koituu kustannuksia yksilölle ja näin käy myös maksuttoman koulutuksen järjestelmissä. Suoria kustannuksia yksilölle ovat materiaaliset kustannukset kuten opiskeluun tarvittavat kirjat ja tietokone. Epäsuorina kustannuksina voidaan nähdä yksilön menettämät ansiotulot opiskelun ajalta.

Suomessa korkeakoulutus on erittäin houkuttelevaa, sillä opiskelua tuetaan verovarojen maksetuilla opintotuilla. Opiskelun aloittaa OECD:n (2007b) mukaan 73 % ikäluokasta, joka on 20 %-yksikköä OECD:n keskiarvoa korkeampi. Opiskelijat maksavat kuitenkin nämä anteliaat tuet takaisin jyrkän progressiivisen verotuksen takia ja opintojen aikaista toimeentuloa tuetaan Suomessa vähiten Pohjoismaista (emt.)

Voidaan sanoa, että pidempi koulutus on yksilön kannalta hyvä sijoitus. Suomessa korkeakoulutuksen hankkineiden osuus kasvanutkin nopeasti, mutta palkkahyöty on samalla heikentynyt kun yksilön markkinahinta on laskenut ylitarjonnan vuoksi (Asplund, 2006, s.123). Lisäksi palkkaetu on kaventunut samalla kun opistoasteen tutkinnon suorittaneiden palkkaetu on vahvistunut, joten sijoituksen kannattavuus on heikentynyt (emt.).

2.2.2 Periaatteellisia näkökulmia korkeakoulutuksen rahoitukseen

Viimeisen kahdekskymmenen vuoden aikana monet kehittyneet maat ovat siirtyneet ajattelemaan, että korkeakoulutus tulisi rahoittaa suoraan siitä hyötäviltä (Curtin, 2000). Näkökulma on syntynyt jo edellä mainitusta ajatuksesta, että korkeakoulutus tuottaa opiskelijalle huomattavaa henkilökohtaista etua.

Kuitenkin maksullisuus ja maksuttomuus -keskustelussa on Henttosen (2008, s.26) mielestä kyse siitä, miten koulutushyvä jakautuu kansalaisten kesken. Tavoitteena on kerätä ja jakaa koulutuksen resurssit tasa-arvoisesti ja oikeudenmukaisesti, mutta näkemyksissä on suuria eroja (emt.).

Maksuttomuuden puolustajien sekä vastustajien kesken on olemassa jonkinlainen konsensus siitä, että koulutuksen tulisi olla tasa-arvon nimissä kaikille mahdollista yksilön taustasta riippumatta (Henttonen, 2008, s.13). Tasa-arvona tässä kontekstissa Henttonen (2008) pitää sitä, ettei yksilön tausta saa vaikuttaa mahdollisuuksiin hankkia koulutus ja hyötyä siitä. Tätä voidaan myös kutsua mahdollisuuksien tasa-arvoksi. Tässä kappaleessa käsittelen siis mahdollisuuksien tasa-arvoa enkä yleisesti tasa-arvoa.

Varsinkin Pohjoismaissa on valloilla käsitys, että maksuton (korkea)koulutus on oikeus, ei etuoikeus, tasa-arvon nimissä. Tämä väite voidaan Barr'in (2004) mielestä kumota ajattelemalla, että jokaisella tulisi olla oikeus ruokaan. Oikeus ruokaan ei tee ruoasta kuitenkaan ilmaista, joten miksi koulutuksen tulisi olla tämän argumentin perusteella ilmaista. Lisäksi korkeakoulutuksesta eivät nauti kaikki kansalaiset, kuten peruskoulusta, jolloin ilmainen koulutus hyödyttää vain tiettyä osaa kansasta. (Barr, 2004.)

Kun puhutaan mahdollisuuksista, on Suomi epätasa-arvoisempi maa koulutukseen pääsyn suhteen kuin esimerkiksi Iso-Britannia ja Hollanti, vaikka näissä maissa on tunnetusti lukukausimaksut käytössä (Usher & Cervenán, 2005). Toisaalta Suomessa 52 % korkeakoulujen opiskelijoista tulee alemmista sosiaalisista luokista (isällä ei korkeakoulutusta), mikä kertoo siitä, ettei Suomessa ole esteitä korkeakoulutuksen hankkimiselle (Asplund, Abdelkarim & Skalli, 2009, s.246). Lisäksi Suomessa 73 % ikäluokasta aloittaa korkeakoulutuksen (yliopistotaso), mikä on paljon OECD:n keskiarvoa 54 % korkeampi (OECD, 2007b). Huolimatta korkeista lukukausimaksuista, myös Australiassa, Yhdysvalloissa sekä Alankomaissa aloitusprosentti on OECD:n keskiarvoa korkeampi. Henttonen (2008, s.32) väittääkin että pelkällä maksuttomuudella ei saavuteta tasa-arvoa. Hänen mukaansa maksuttomuuden lisäksi on myös oltava olemassa tukimenetelmiä, kuten Suomessa käytössä oleva opintotuki, jotta pääsy koulutukseen piiriin olisi tasa-arvoista.

OECD:n (2007b) raportin mukaan taas maat, joissa PISA –tutkimuksen mukaan perusasteen koulujen välillä oli vain vähän eroa, sijoittuivat myös korkealle korkeakoulujen tasa-arvoisessa sisäänpääsyssä. Korrelaatio ei kuitenkaan kerro kausaalisuhteesta.

Toinen epäkohta Henttonen (2008) mielestä suomalaisen järjestelmässä tasa-arvon kannalta on se, että korkeakoulutusta tuetaan nyt samoin kuin julkista terveydenhuoltoa. Terveys- ja terveydenhuoltohan on olemassa turvatakseen tappiot, jotka yksilölle tuottavat epäonnin tai sairaus. Samalla periaatteella yksilö on koulutukseen osallistumisen varalta vakuutettu nykyisessä mallissa. Tämä on ristiriidassa, sillä koulutushan tuo pelkkää hyvää yksilölle pitkällä tähtäimellä ja etuus koskee vain osaa kansasta. (Henttonen, 2008, s.14)

Maksuttomuus ei näytä tuottavan automaattisesti tasa-arvoa ja maksuttomuus ei ole välttämättä tehokkain tapa avata koulutus kaikille ja jakaa koulutuksen osallistumisen riskit. Kuitenkin maksuttomuus toimii nyt Suomessa hyvänä keinona yhteiskunnan tasa-arvoistamisessa, ja yhteiskunta saa käyttämänsä resurssit takaisin tulevaisuudessa korkeammista verotuloista. (Henttonen, 2008, s.26)

Uutena ulottuvuutena keskustelussa ovat ulkomaalaisten opiskelijoiden kasvava määrä korkeakouluissa. Pitäisikö heidänkin olla tasa-arvoisessa asemassa koulutukseen pääsyn suhteen, jos he eivät koulutuksesta maksa edes verotulojen muodossa?

Lukukausimaksujen käyttöönoton ja/tai korotuksen vaikutukset koulutukseen hakeutuviin

Argumenttina maksuja vastaan käytetään usein maksujen käyttöönoton tuomia negatiivisia vaikutuksia koulutuksen kysyntään. Tutkimusten mukaan (esim. Chapman & Ryan, 2005) opiskelijat ovat ryhmänä hyvin joustamattomia opiskelun hinnanmuutosten suhteen eli maksujen nosto tai lasku ei vaikuta merkittävästi opiskelemaan hakevien määrään. On kuitenkin todisteita siitä, että maksujen nousu vaikuttaa juuri alemmista sosioekonomisista luokista tulevien opiskelijoiden määrään negatiivisesti (Marcucci & Johnstone, 2007). Näin kuitenkin ei tapahtunut maissa, joissa maksut voi maksaa myös valmistumisen jälkeen (emt.). Myös Coelli (2009, s.1095) painottaa maksujen esittämisellä olevan huomattava negatiivinen vaikutus nimenomaan alempien sosiaalisten luokkien suhteelliseen osuuteen opiskelijoista. Negatiivinen vaikutus ei ole tutkimuksen mukaan merkittävä keski- ja hyvätuloisten joukossa.

Tanskassa ja Ruotsissa lukukausimaksujen käyttöönotto on johtanut hakijamäärien romhdukseen etenkin Afrikan maista, mutta hakijamäärien oletetaan tasaantuvan pitkällä aikavälillä (Ulkoministeriön kehitysviestintä, 2011).

Ilmainen korkeakoulutus on tulonsiirto köyhiltä rikkaille

Muun muassa Barr, Chapman ja Ryan sekä Johnstone (2004; 2005; 2004) ovat sitä mieltä, että yleisistä verovaroista tuettava korkeakoulutus on tulonsiirto köyhiltä rikkaille. Verorahoin tuettavassa mallissa kaikki maksavat korkeakoulutuksesta huolimatta siitä, osallistuuko koulutukseen vai ei. Epätasa-arvoisen mallista tekee sen, että koulutus on hyvin periytyvää eli koulutettujen lapsia on suhteellisesti enemmän korkeakouluissa kuin alemman sosioekonomisen tason taustaisia lapsia. Tällöin huonompioisaiset maksavat hyväosaisten koulutuksesta sekä auttavat heitä ansaitsemaan yhä paremmin. (Chapman & Ryan, 2005; Johnstone, 2004.)

Curtin (2000) pitää edellä esitettyjä väitteitä perättöminä. Köyhät ja kouluttamattomat maksavat usein vähemmän veroja, mitä ehtivät elinaikanaan kuluttaa esimerkiksi sosiaalipalveluihin, jolloin heidän panoksensa muiden korkeakoulutukseen olisi olematon. Korkeakouluttautuneet ansaitsevat paremmin ja siten maksavat enemmän veroja. Näin he kattavat oman koulutuksensa sekä käyttämiensä sosiaalipalvelujen kustannukset sekä osan köyhempien saamista tuista. Korkeakoulutus on yhteydessä korkeampaan tulotason, ja siten lisäksi suurempaan veroprosenttiin useimmissa progressiivisen verotuksen maissa. Myös kulutus on suurempaa enemmän ansaitsevilla ja he maksavat enemmän veroja myös arvonlisäveron muodossa. (Curtin, 2000.)

Barr (2004) on sitä mieltä, että korkeasti koulutetut maksavat tuloveroissaan moninkertaisesti koulutukseen käytetyt rahat, ei ole totta. Englannissa tuloverotuksen tuomat varat muodostavat valtion budjetista vain 25 % (Suomessa vuonna 2010 n. 20 % ²) ja vuonna 2001 työntekijöistä suurimmalla osalla 82 % ei ole korkeakoulutusta (Suomi 66 % ³). (Barr, 2004.)

²http://www.stat.fi/tup/suoluk/suoluk_valtiontalous.html

³http://www.stat.fi/til/tyti/2010/16/tyti_2010_16_2011-11-01_tau_009_fi.html

Maksut toimivat kannustimena

Curtinin (2000) mukaan lukukausimaksuja on perusteltu myös opiskelijoiden motivoimisella innokkaampaan, tehokkaampaan opiskeluun sekä nopeaan valmistumiseen. On kuitenkin osoitettu, että opiskelijoiden opiskelumotivaatioon vaikuttavat enemmän tunnustukset hyvien arvosanojen vuoksi saaduista palkkioista. Maksut olisivat siis rangastus myös ahkerille ja nopeasti valmistuville opiskelijoille. (Curtin, 2000.)

2.2.3 Korkeakoulutuksen rahoitusmalleja

Koulutuksen rahoituksen valtionosuuden väheneminen yhdessä muiden tekijöiden kanssa on johtanut viimeisen 25 vuoden aikana siihen, että useimmat Euroopan maat ovat siirtyneet jonkinlaiseen maksullisuuteen korkeakouluissaan (Hölttä, 1998). Myös Pohjoismaissa, joissa maksuton koulutus on perinteisesti mielletty enemmänkin oikeudeksi kuin etuoikeudeksi, on siirrytty pikkuhiljaa lukukausimaksujen käyttöönottoon, ensimmäiseksi EU- ja ETA -maiden ulkopuolelta tuleville opiskelijoille.

Korkeakoulutuksen rahoituksen eri mallit liittyvät vahvasti ajatukseen vanhempien vastuusta osallistua rahallisesti lapsiensa koulutukseen. Niissä maissa, joissa lukukausimaksut ovat käytössä (esim. USA, Itävalta, Alankomaat) oletetaan vanhempien ottavan vastuun lapsensa opiskelujen kuluista. Pohjoismaissa, jossa koulutus on ollut pisimpään maksutonta, sekä Australiassa ja Skotlannissa, joissa koulutuksen kulut maksetaan valmistumisen jälkeen, ajatellaan, ettei opiskelijoiden tai heidän vanhempiensa tulisi maksaa koulutuksesta ainakaan koulutuksen aikana. (Marcucci & Johnstone, 2007.)

Seuraavaksi kerron lyhyesti erilaisista korkeakoulutuksen rahoitusmalleista, jotka ovat parhaillaan käytössä lähinnä eri länsimaissa. Näissä malleissa ei ole otettu huomioon opintojen tukea vaan ne keskittyvät ainoastaan opintojen maksullisuuteen. Mallit voidaan jakaa neljään eri luokkaan: maksuton koulutus, lykätty maksu, kaksoisväylä sekä perinteiset lukukausimaksut. Tarkastelen laajemmin Suomea ja suomalaisten korkeakoulujen tilannetta kohdassa 2.3.

Maksuttomuus

Täysin maksuton koulutus on enää harvinaista ja käytössä vain Pohjoismaissa sekä joissain Saksan liittovaltioissa. Tässä mallissa valtiot tukevat verorahoista kerätyillä varoilla korkeakoulutusta hyvin anteliaasti. Suomessa tuki on erityisen laaja, sillä valtio tukee sekä suoraan koulutusinstituutioita sekä lisäksi opiskelijoita opintotuen sekä muiden tukien, kuten ateriatuen, muodossa. Mallin hyvyttenä on pidetty sitä, että maksukykyyn liittyviä esteitä opinnoille ei ole, mikä johtaa tai sen oletetaan johtavan tasa-arvoiseen, tulotasosta riippumattomaan, pääsyyn korkea-asteen opetuksen piiriin. Tulotasosta riippumaton sisäänpääsy on toteutunut hyvin Suomessa, sillä 52 % opiskelijoista tulee alemmista sosiaaliluokista (Asplund, 2006). On kuitenkin otettava huomioon, ettei pelkkä maksuttomuus välttämättä aiheuta mahdollisuuksien tasa-arvoa vaan vaikutus muodostuu yhdessä opiskelijoiden tukien kanssa (Marcucci & Johnstone, 2007).

Lykätty maksu

Muun muassa Australiassa ja Skotlannissa on käytössä järjestelmä, jossa lukukausimaksut voi halutessaan maksaa opintojen jälkeen. Opiskelija ottaa käytännössä tulevaan tu-

lotasoon sitoutettua lainaa, jonka hän maksaa valmistuttuaan takaisin kykyjensä mukaan.

Australiassa opiskelijoiden vanhemmilla on mahdollisuus maksaa lukukausimaksut heti lapsensa opiskeluaikana tai jättää ne lapsensa maksettavaksi tuloista riippuvan lainan muodossa opiskelujen jälkeen (lykätty maksu). Takaisinmaksu perustuu siis tuleviin ansioihin progressiivisen verotuksen tavoin. Mitä enemmän opiskelija ansaitsee valmistumisensa jälkeen, sitä enemmän hän maksaa koulutuksestaan, mutta vain tiettyyn summaan asti. Jos ansiotulot jostain syystä eivät ylitä keskipalkkaa koskaan sairastumisen tai muun syyn vuoksi, opiskelija ei joudu maksamaan koulutuksesta mitään. (Marcucci & Johnstone, 2007.)

Kaksoisväylä

Lainopillisten esteiden tai vahvan julkisen vastustuksen vuoksi joissakin maissa on otettu käyttöön niin kutsuttu ”kaksoisväylä” (*dual-track*) malli. Tässä mallissa osa kolmannen asteen koulutuspaikoista on maksuttomia tai lähes maksuttomia esimerkiksi toisen asteen loppukokeissa parhaiten pärjänneille. Lopuille päteville opiskelijoille koulutus on maksullista. (Marcucci & Johnstone, 2007.)

Toinen esimerkki kaksoisväylä-mallista on Tanskassa ja Ruotsissa käyttöönotetut maksut maisteriohjelmissa EU- ja ETA-maiden ulkopuolisille opiskelijoille. Näissä maissa maan omat sekä EU- ja ETA-maiden kansalaiset ovat oikeutettuja ilmaiseen korkeakoulutukseen ulkopuolelta tulleiden opiskelijoiden maksaessa lukukausimaksuja.

Up-front eli etukäteen maksettavat lukukausimaksut

Etukäteen perittävien lukukausimaksujen mallissa oletetaan, että opiskelija itse tai hänen perheensä osallistuu koulutuksen kustannuksiin jo koulutuksen aikana (Marcucci & Johnstone, 2007). Vähävaraisille on usein olemassa ohjelmia ja tukia kuten stipendejä, joilla helpotetaan koulutuksen tuomaa rahallista taakkaa. Näin jokainen osallistuu periaatteessa koulutuksen kustannuksiin oman kykynsä mukaan. Keskituloisissa perheissä, etenkin USA:ssa, lapsen koulutusta varten säästäminen aloitetaan jo lapsen syntymästä lähtien. Malliin liittyy selkeästi koulujen vahva profiloituminen sekä joissain maissa (esim. USA) maksujen suuri variaatio koulun maineen, ohjelman suosion sekä koulutuksen tuottaman palkkapreemion eli tutkintojen antamien tuottojen johdosta (Vossensteyn, 2009).

2.3 Lukukausimaksukokeilu Suomessa

Hiljattain myös Suomessa taloudelliset alamäen jaksot, kasvava koulutuksen kysyntä sekä väestörakenteen muutokset ovat johtaneet toimiin rahoituspohjan laajentamiseksi (Hölttä, 1998). Lisää kierroksia keskusteluun ovat tuoneet nykyaikainen tehokkuusajattelu sekä kasvavat paineet Euroopan Unionin ja OECD:n suunnalta (Euroopan komissio, 2011; OECD, 2007b).

Seuraavaksi esittelen tekijöitä, jotka ovat Suomessa johtaneet lukukausimaksukokeilun käyttöönottoon ja sitä millaisilla argumenteilla kokeilua on Suomessa puolustettu sekä kritisoitu.

2.3.1 Lukukausimaksukokeiluun johtaneita tekijöitä

Muiden eurooppalaisten maiden tavoin on Suomessa 1980-luvulta lähtien siirrytty korkeakoulutuksen ohjauksessa ja rahoituksessa suorasta valtion kontrollista kohti korkeakoulujen suurempaa päätäntävaltaa ja toiminnan tuloksellisuutta korostavaa ohjausta (Hölttä, 1998). 1990-luvun alun taloudellinen lama toi lisää paineita yliopistojen rahoituspohjan uudistamiseen ja valtion budjettien kiristyessä korkeakoulujen oli jatkettava toiminnan taloudellisen tehokkuuden korostamista sekä etsittävä uusia rahoituslähteitä (emt.). Rahoituksen supistumisesta huolimatta korkeakoulutus ja tutkimus nähtiin tärkeänä osana Suomen taloudellisen kehityksen aikaansaamisessa ja se kytkettiin osaksi talous- ja innovaatiopolitiikkaa (Puustelli, Niskanen, Sund & Rajala, 2009, s.10). Vuoden 2008 lopulla talous ajautui jälleen taantumaan ja säästöpainet tulivat uudestaan esille.

Koulutuksen uudistamiseen ja rahoituspohjan laajentamiseen on tuonut lisäpainetta yhtenäisen EU:n strategia ja kasvava kilpailu Amerikan ja Aasian kanssa. Euroopan komissio (2011) on Lissabonin strategiaa seuraten peräänkuuluttanut opinto-ohjelmien nykyaikaistamista sekä korkeakoulutuksen tehokkaampaa rahoitusta ja hallintotapaa. Tavoitteena on siirtyä maailmanlaajuisen osaamistalouteen, kasvua ja työllisyyttä edistäen. Komission (2011) mukaan Euroopan yliopistojen potentiaalia ei ole käytetty täysin hyväksi usein puutteellisen hallinnointitaidon ja riittämättömän rahoituksen vuoksi. Haasteena on luoda yhtenäiset ja houkuttelevat koulutusmarkkinat. Korkeakoulut edelleen aliarvioivat niitä etuja, joita tietojen jakamisesta talouselämän ja yhteiskunnan kanssa voitaisiin saada, eivätkä yritykset toisaalta kykene riittävästi omaksumaan uutta tietoa, jotta ne voisivat valjastaa korkeakoulujen tutkimustyön tarjoamat mahdollisuudet omaan käyttöönsä. (Euroopan komissio, 2011)

Tästä johtuen myös Suomen hallituksen *Koulutuksen ja tutkimuksen kehittämissuunnitelma vuosille 2007-2012* (2008) - raportin mukaan koulutus- ja tutkimusjärjestelmää kehitetään kokonaisuutena siten, että globalisaation ja kansainvälistymisen haasteisiin, väestörakenteen, työelämän ja muihin toimintaympäristön muutoksiin vastataan niin, että Suomen kilpailukyky kasvaa, hyvinvointi lisääntyy ja niin, että kulttuuri, luovuus ja sivistys vahvistuvat.

Suomessa, niin kuin muissakin länsimaissa, väestöllisen huoltosuhteen kehitys on huolestuttavaa. Opetusministeriön (2010) mukaan talouskasvun ja hyvinvoinnin takaminen edellyttää työikäisen väestön työllisyysasteen kasvattamista. Raportin mukaan työuria on onnistuttu pidentämään onnistuneesti loppu- muttei alkupäästä. Näin lukukausimaksuja on esitetty Suomessa keppinä, jotta opiskeluaikoja saataisiin lyhennettyä, sillä OECD:n (2007b) mukaan suomalaiset ovat keskimääräistä vanhempia valmistuessaan. Suuren osan tästä selittää se, että Suomessa valmistutaan pääasiassa maisteriksi viisi vuotta kestävästä koulutuksesta kun taas esimerkiksi USA:ssa valmistutaan pääasiassa kandidaattitasolla (Henttonen, 2008, s.19). Toinen selittävä tekijä, on se ettei Suomessa opiskelemaan päästä suoraan lukiosta. Opintoaikoja on pyritty jo kertaalleen lyhentämään 1990-luvulla ottamalla käyttöön opintotuet pelkkien lainojen sijaan, mutta haluttua vaikutusta ei kuitenkaan ole nähty (Häkkinen & Uusitalo, 2003).

Lisäksi lukukausimaksujen puolesta puhuu korkeakoulutuksen voimakas periytyminen Suomessa (OECD, 2007a, s. 116–122). Koulutetuilla vanhemmilla on kyky auttaa ja tukea lasta opintiellä ja he osaavat myös vaatia lapselta hyvää koulumenestystä.

Henttonen (2008, s.37) huomauttaa vielä, että Suomessa eletään vanhentuneen, sul-

jetun järjestelmän logiikan mukaan. Opiskelun ja työmarkkinoiden kansainvälistyessä myös liikkuvuus on lisääntynyt ja koulutuksen kustannuksiin osallistuminen voidaan välttää vaikka etuudesta olisi nauttinut. Yhtenä ratkaisuna tähän ongelmaan olisivat juuri lukukausimaksut. Toisaalta yksilö ei pääse helposti näkemään koulutuksen kannattavuutta Suomessa, sillä taloudellinen hyöty on verrattain pieni. Tämä asettaa haasteita suomalaisen työmarkkinoiden houkuttelevuudelle, jos haluamme pitää lahjakkaat Suomessa ja lisäksi houkutellessa työntekijöitä tänne. (Henttonen, 2008.)

2.3.2 Kokeilun toteutus

Monien muiden EU-maiden esimerkkiä seuraten opetusministeriö asetti alkuvuodesta 2005 työryhmän selvittämään EU/ETA-maiden ulkopuolelta tulevien tutkinto-opiskelijoiden koulutuksesta aiheutuvien kustannusten kattamista tulevaisuudessa maksuilla (Opetusministeriö, 2005). Työryhmä esitti elokuussa 2005 lukukausimaksujen käyttöönottoa EU- ja ETA-maiden ulkopuolelta tuleville korkeakouluopiskelijoille tutkintoon johtavassa koulutuksessa. Yleisistä lukukausimaksuista kuitenkin luovuttiin ja Matti Vanhasen toisen hallituksen ohjelmaan (2007) kirjattiin lause laadukkaasta ja maksuttomasta koulutuksesta hyvinvointiyhteiskunnan perustana. Edellä mainitusta dokumentista löytyy kuitenkin myös lause, jolla mahdollistetaan maksukokeilu yksittäisissä maisteriohjelmassa EU- ja ETA-maiden ulkopuolisille opiskelijoille. Lukukausimaksukokeilu määritettiin ajalle 1.1.2010 – 31.12.2014 ja se on vapaaehtoinen korkeakouluille (yli, 2009, 10 §). Maksukokeilun edellytyksenä on stipendijärjestelmä, joka tulisi olla suunnattu vähävaraisille opiskelijoille.

Yliopistolaki 10 §, Koulutusohjelman maksullisuus:

Yliopisto voi periä maksuja vieraskieliseen ylempään korkeakoulututkintoon johtavaan koulutusohjelmaan hyväksytyiltä. Maksujen perimisen edellytyksenä on, että yliopistolla on apurahajärjestelmä, jolla voidaan tarvittaessa tukea maksulliseen maisteriohjelmaan osallistuvien opiskelua. Maksua ei voida kuitenkaan periä Euroopan talousalueeseen kuuluvan valtion kansalaiselta ikä siltä, joka Euroopan yhteisön lainsäädännön tai Euroopan yhteisön ja sen jäsenvaltioiden muun sopimuspuolen kanssa tekemän sopimuksen mukaan rinnastetaan Euroopan unionin kansalaiseen.

Asetus: Sivistysvaliokunnan mietintö SiVM 5/2009 vp:

Kokeilun edellytyksenä on ulkomaisille vähävaraisille opiskelijoille tarkoitettu stipendijärjestelmä.

Opetus- ja kulttuuriministeriö (2011) asetti työryhmän arvioimaan lukukausimaksukokeilua. Siihen kuuluu jäseniä opetuksen erialoilta ja kokeiluun osallistuvista kouluisista. Työryhmän tehtävänä on selvittää miten maksullisten koulutusohjelmien olemassa olo vaikuttaa korkeakoulujen kansainvälistymiseen, suomalaisen korkeakoulutuksen vetovoimaan, opiskelijavirtoihin ja vieraskielisen korkeakoulutuksen laatuun. Arvioinnin tavoitteena on tuoda esille lukukausimaksukokeilun tuomia muutoksia verrattuna korkeakouluissa maksuttomasti toteutettuun vieraskieliseen koulutukseen (Lukukausimaksukokeilun seuranta- ja arviointityöryhmä, 2011).

Kokeilu on vapaaehtoinen korkeakouluille. Halukkaiden koulujen oli ilmoitettava kesäkuuhun 2011 mennessä miten ne aikoivat kokeiluun mukaan lähteä. Maksukokeilun

arviointityöryhmän (2011) väliraportin mukaan kokeiluun oli ilmoitettu raportin valmistumiseen 21.4.2011 mennessä 124 englanninkielistä maisteriohjelmaa yhdeksästä yliopistosta ja 18 YAMK-ohjelmaa kymmenestä ammattikorkeakoulusta. Näissä mukana ovat jo käynnissä olevat EU:n koordinoimat Erasmus Mundus -ohjelmat.

Kyselyn valmistuessa suurimmalta osalta kokeiluun ilmoitetuista ohjelmista puuttui edelleen päätös lukukausimaksujen perimisen aloittamisvuodesta. Aalto- ja Lappeenrantaan teknillinen yliopisto aloittivat kokeilun vuoden 2011 syksyllä. Molemmilla yliopistoilla lukuvuoden hinta on 7 000 tai 8 000 euroa.

Yleisesti kokeiluun lähdön syiksi listattiin kansainvälistymisen tärkeys sekä korkeakoulun profiloituminen. Kokeilu nähtiin keinona kehittää ja testata kansainvälisen koulutuksen laatua sekä vetovoimaa. Lukukausimaksukokeilu koettiin korkeakouluissa osaksi koulutusviennin kokonaisuutta ja mahdollisuudeksi pilotoida koulutusliiketoimintaa. Kokemuksia haluttiin kerätä myös apurahajärjestelmistä. Käytännönläheisimpinä perusteluina esitettiin myös kansainvälinen kehitys ja muiden maiden, etenkin Ruotsin, esimerkki. Koulujen olettamuksena oli, että EU:n ulkopuolisilta perittävät lukukausimaksut tulevat myöhemmin käyttöön myös Suomessa, jolloin kokeilu tarjosi mahdollisuuden harjoitella tätä. (Lukukausimaksukokeilun seuranta- ja arviointityöryhmä, 2011.)

2.3.3 Kokeilua kohtaan esitetty kritiikki

Lukukausimaksukokeilua ovat kritisoineet niin yliopistojen johto kuin ylioppilaskunnat. Huolenaiheena on ollut maksujen esittämisen jälkeinen hakijamäärien putoaminen, kuten tapahtui Tanskassa ja Ruotsissa. Suomessa vain osa maisteriohjelmista lähti mukaan kokeiluun, joten maksut koskevat hyvin pientä joukkoa, eikä hakijamäärien muutoksilla ole suurta vaikutusta. Aalto-yliopistossa hakijamäärät eivät laskeneet huomattavasti kaupallisilla tai taide-aloilla, mutta kun ei oteta huomioon 2011 alkanutta uutta ohjelmaa, teknillisen alan hakijamäärät laskivat 39 % (Sauli, 2011).

Aalto-yliopiston ylioppilaskunnan AYY:n pelkona on että eriarvoisuus kasvaa maksavien ja ilmaiseksi opiskelevien joukossa. Jo nyt Aalto-yliopisto tarjoaa maksukokeiluohjelmien opiskelijoille intensiivisen suomen kielen kurssin ennen opintojen alkua, mitä ei ole tarjolla muille opiskelijoille. Lisäksi maksuohjelmille on annettu kullekin 25 000 euroa kehittämisrahaa, josta hyötyvät myös kyseisissä ohjelmissa ilmaiseksi opiskelevat (Sauli, 2011).

Toisena AYY:n pelkona on, että maksut estävät lahjakkaiden vähävaraisten opiskelijoiden tulon Suomeen ja opiskelija-aineksen laatu kärsii. On kuitenkin huomattava, että maisteriohjelmaan liittyvät vaatimukset, mm. alempi korkeakoulututkinto, edellyttävät jo investointeja koulutukseen, eivätkä Suomeen ja Aalto-yliopistoon tulevat opiskelijat ole koskaan olleet kaikkein köyhimmästä päästä. (emt.)

Ylioppilaskunnat pelkäävät myös porttiteoriaa eli että kokeilu olisi vain välivaihe kaikille maksullisen koulutuksen tiellä (Sauli, 2011). Lukukausimaksukokeilun seuranta- ja arviointityöryhmän väliraportin (2011) mukaan korkeakouluille suurin syy maksukokeiluun mukaan lähtemiseen oli juuri se, että maksuja pääsisi kokeilemaan ennen niiden laajempaa käyttöönottoa.

Yle Uutisten mukaan maksukokeilu oli monille korkeakouluille liian lyhyt (YLE, 19.5.2011). Uutisen mukaan mm. Helsingin ja Itä-Suomen yliopisto eivät osallistu 2011 – 2014 kestäväan kokeiluun sen lyhyiden vuoksi, sillä yliopistojen johdon mielestä sekä opiskelijamäärät että kokemus jäisivät vähäisiksi.

Etenkin SYL:n (Suomen ylioppilaiden liitto) huolena on ollut myös se, että ulkoministeriö suunnittelee kansallista apurahaohjelmaa kehitysmaista Suomeen tuleville opiskelijoille, ja kansainvälisten kriteerien mukaan sen kulut voidaan laskea kehitysavuksi. Liiton mukaan koulutus on tärkeä osa kehitysyhteistyötä, mutta kehitysapua vain jos se toteutetaan avunsaajamaiden tarpeista lähtien ja niiden kanssa yhteistyössä. Hänen mielestään kehitysavun tulee tähdätä ensisijaisesti kehitysmaiden talouden ja hyvinvoinnin edistämiseen. (Kepa ry, 3.2.2011)

Lisäksi AYY (2010) on esittänyt huolensa lukukausimaksukokeilun vaikutusten asianmukaisesta arvionnista. Seuranta- ja arviointityöryhmä ei oman suunnitelmansa (2010) mukaan arvioi esimerkiksi vaikutuksia opiskelija-ainekseen.

2.3.4 Lukukausimaksukokeilu Aalto-yliopistossa

Vararehtori Martti Raevara ja Hannu Seristö perustivat maaliskuussa 2010 työryhmän valmistelemaan sekä ideoimaan tapaa, jolla Aalto-yliopiston tulisi ottaa osaa opetusministeriön valmistelemaan lukukausimaksukokeiluun. Aalto-yliopiston tavoitteena oli osallistua kokeiluun jo syksyllä 2011 alkavana lukukautena, jotta kokeilun vaikutuksista ja käytännöistä saataisiin mahdollisimman paljon tietoa myöhempiä päätöksiä varten.

Työryhmään kuului jäseniä kaikista Aalto-yliopiston korkeakouluista mahdollisimman hyvän tuloksen varmistamiseksi. Työryhmässä oli myös opiskelijaedustus Aalto-yliopiston ylioppilaskunnan AYY:n puolesta. Työryhmän tavoitteena oli etsiä parhaimmat ratkaisut osallistuvien ohjelmien, maksutasojen, stipendijärjestelmän sekä hakuprosessin kannalta. Työryhmä konsultoi kaikkien Aalto-yliopiston koulujen ohjelma-, laitos- ja tiedekuntajohtajia kahdesti ja teetti vertailevan tutkimuksen yksityisellä konsulttitoimistolla lukukausimaksujen yleisestä tasosta kansainvälisesti. (Aalto Working Group, 2010.)

Työryhmä pohtii loppuraportissaan (emt.) eri vaihtoehtojen positiivisia ja negatiivisia puolia, mutta ei esitä mitään konkreettista ratkaisua tai yhtä parasta mallia. Raportin yhteenvedossa työryhmä muistuttaa, että päätös kerätä maksuja eroaa päätöksestä kerätä maksua vain osasta ohjelmista. Työryhmä myös painottaa huomiotaan, jonka mukaan (korkeampi) hinta on usein merkki (korkeammasta) laadusta. Aalto-yliopiston tulisi hinnoitella kaikki maisteriohjelmansa samoin, jotta opiskelijalle ei synny kuvaa eritasoisista ohjelmista. Huono opetuksen laatu on joka tapauksessa haitallista Aalto-yliopiston brändille huolimatta siitä peritäänkö maksuja vai ei.

Ylioppilaskunnan ajamat asiat

AYY vaatii linjapaperissaan (2010), että lukukausimaksukokeilun hyödyt tulee osoittaa. AYY ilmaisi olevansa huolissaan yliopistojen kansainvälisyydestä sekä monimuotoisuudesta, sillä maksujen myötä opiskelijat saattavat kadota Suomen oppilaitoksista. Näin on käynyt esimerkiksi Tanskassa ja Ruotsissa, jossa vastaavat lukukausimaksut on otettu käyttöön. AYY on paperin mukaan skeptinen lukukausimaksujen käyttöönotosta koituvien hyötyjen suhteen ja haluaa muistuttaa, että opiskelijat eivät ole pelkkä menoerä vaan kansainvälisyys on yksi huippuyliopiston perusedellytyksistä.

AYY (2010) painottaa, että lukukausimaksua on arvioitava monipuolisesti ja läpinäkyvästi. Lisäksi kokeilun tulisi olla mahdollisimman todenmukainen, jotta arvioinneista tai koko kokeilusta olisi hyötyä. Myös stipendijärjestelmien tarveharkintaisuus

on AYY:n mielestä tärkeä asia. Näin taataan myös vähempiosaisille opiskelijoille mahdollisuus vastaanottaa opiskelupaikka. AYY muistuttaa lisäksi opiskelijoiden yhdenvertaisuudesta, joten kaikille opiskelijoille on taattava samat laadukkaat palvelut maksuista riippumatta.

AYY tarkensi linjaansa jyrkemmäksi AYY:n hallitusohjelmataavoitteissa vuosille 2011–2015. Niissä vaaditaan tutkintoon johtavan koulutuksen säilyvän maksuttomana kaikille opiskelijoille. (Henkilökohtainen viestintä, kevät 2011)

Toteutus

Aalto-yliopisto lähti mukaan lukukausimaksukokeiluun rehtori Tuula Teerin päätöksellä (2010). Maksukokeiluun otettiin yhteensä 14 englanninkielistä maisteriohjelmää (ks. taulukko 2.2) mukaan lukien EU:n koordinoimat Erasmus-mundus -ohjelmat, joita tässä työssä ei käsitellä tarkemmin. Teknillisestä korkeakoulusta valittiin mukaan kokeiluun neljä ohjelmaa, kauppa- ja korkeakoulusta kolme, taideteollisesta korkeakoulusta yksi ohjelma sekä jokaisesta koulusta IDBM-ohjelma. Kaikkien ohjelmien lukukausimaksujen suuruudeksi määrättiin 8 000 euroa lukuvuodessa. Kaksivuotisten maksullisten maisteriohjelmien kokonaishinnaksi tuli siis 16 000 euroa (vrt. esim. London School of Economics £ 21 000 eli 25 000 €/vuosi, Ecole Polytechnique 11 000 €/tutkinto)

Taulukko 2.2

Ohjelman nimi	Kelpoiset hakijat		Valitut		Vastaanotetut paikat	
	Kaikki	ei EU	Kaikki	ei EU	Kaikki	ei EU
Communications Ecosystems	25	14	12	10	8	7
Environmental Art	15	5	7	3	6	2
IDBM/HSE	36	1	12	1	11	1
IDBM/TAIK	58	26	12	8	7	3
IDBM/TKK	25	13	12	7	9	4
Information and Service Management	61	5	43	4	34	2
International Business Communication	42	2	23	2	23	2
Marketing	8	2	8	2	7	2
Mechanical Engineering	67	31	32	24	15	8
Mobile Computing	63	29	20	11	7	5
Service Design and Engineering	79	21	24	9	11	4
Yhteensä	479	149	205	81	138	40

Lähde: Sanna Heikkinen, Planning Officer/Scholarship Programmes, Aalto-yliopisto

Kuten uusi yliopistolaki vaatii, on lukukausimaksujen käyttöönoton vaatimuksena stipendijärjestelmä (yliopistolaki, 10 §). Matti Vanhasen toisessa hallitusohjelmassa (2007) ohjataan stipendijärjestelmän käyttämistä ”ulkomaisille vähävaraisille opiskelijoille”. Aalto-yliopisto päätti perustaa stipendijärjestelmän akateemisiin ansioihin eli samoihin kriteereihin, joilla opiskelijat tulivat valituksi sisään.

Aalto-yliopiston stipendijärjestelmässä on neljä eri luokkaa (taulukko 2.3). Aalto-yliopisto maksaa A-kategorian opiskelijoiden puolesta ohjelman lukukausimaksut ja lisäksi he saavat tukea elinkustannuksiin 8 000 euron verran lukuvuodessa. B-kategorian opiskelijoiden puolesta maksetaan lukukausimaksut, mutta ei ylimääräisiä kuluja. C-kategoriassa opiskelijat saavat alennuksen lukukausimaksuista ja joutuvat maksamaan vain puolet summasta. D-kategorian opiskelijat maksavat täydet lukukausimaksut.

Opetusministeriön työryhmän teettämän kyselyn mukaan (2010) Aalto-yliopisto aikoo jakaa opiskelijat opintomenestyksen perusteella niin, että n. 10 % opiskelijoista kuuluisi kategoriaan A, 25 % kategoriaan B, 35 % kategoriaan C ja loput 30 % kategoriaan D kuten taulukosta 2.3 ilmenee.

Taulukko 2.3. Apurahakategoriat Aalto-yliopistossa

Stipendikategoria	Kattavuus	€/ vuosi	Osuus opiskelijoista
A	Kaikki maksut ja elinkustannukset	16 000	10 %
B	Kaikki maksut	8 000	25 %
C	Puolet maksuista	4 000	35 %
D	Ei stipendiä	0	30 %

Lähde: Lukukausimaksukokeilun seuranta ja arviointi työryhmän väliraportti I (2010).

Maksukokeilussa mukana oleviin maisteriohjelmiin saapui yhteensä 149 tarkastuksen läpäissyttä hakemusta EU- ja ETA maiden ulkopuolisilta hakijoilta (taulukko 2.2). Hakemuksista 108 tuli teknillisen alan ohjelmiin, 10 kauppakorkeakouluun ja 31 taideollisen alan ohjelmiin. Aalto-yliopiston maksukokeilussa mukana oleviin maisteriohjelmiin hyväksyttiin yhteensä 81 ulkoeurooppalaista opiskelijaa, joista A-kategorian apurahan sai kahdeksan opiskelijaa (10 %), B-kategorian apurahan 27 opiskelijaa (33 %) ja C-kategorian apurahan 32 opiskelijaa (39 %) ja 15 opiskelijaa sijoittui kategoriaan D, jolloin heille tulee maksettavaksi täydet lukukausimaksut eli 8000 euroa lukuvuodessa kuten taulukosta 2.4 nähdään. Tilanne paikan vastaanottaneiden varmistuttua näkyy myös taulukosta 2.4.

Taulukko 2.4. Stipendien jakautuminen

Kategoria	Suunniteltu osuus %	Hyväksytyt		Paikan vastaanottaneet		Tuotto
		Lkm	%	Lkm	%	
A	10 %	8	10 %	8	20 %	- 64 000
B	25 %	27	33 %	20	50 %	0
C	35 %	32	32 %	9	23 %	36 000
D	30 %	14	17 %	3	8 %	24 000
Yhteensä	100 %	81	92 %	40	100 %	- 4 000

Lähde: Sanna Heikkinen, Planning Officer/Scholarship Programmes, Aalto-yliopisto

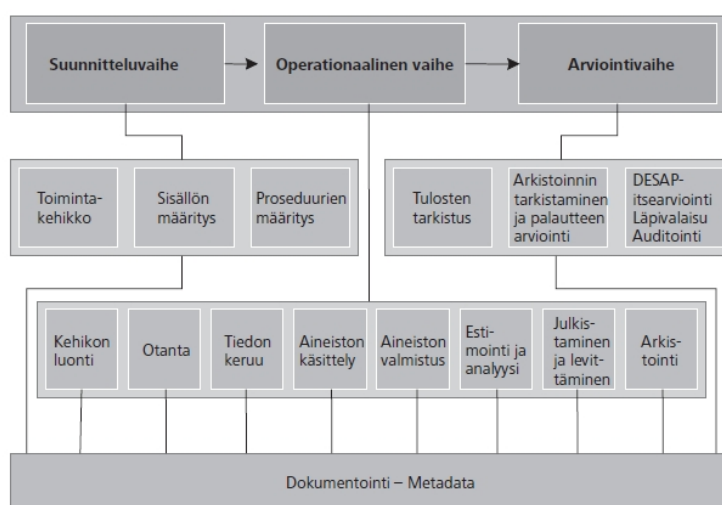
Ainakin tällä hetkellä kokeilu tuottaa tappiota Aalto-yliopistolle (taulukko 2.4). Paikan vastaanottaneiden perusteella Aalto-yliopisto saa opiskelijoilta maksuista yhteensä 60 000 euroa ja tulee maksamaan opiskelijoille 64 000 euroa stipendien muodossa. Opiskelijoille on lisäksi kustannettu intensiivikielikurssi ennen lukukauden alkamista ja maksukokeilussaoleville ohjelmille on annettu ylimääräistä kehittämisrahaa (henkilökohtainen viestintä, toukokuu 2011).

3 Kyselytutkimuksen suunnittelu ja toteutus

Kyselytutkimuksella tarkoitetaan tutkimusta, jossa tutkija esittää vastaajalle kysymyksiä kyselylomakkeen välityksellä. Kyselytutkimusta pidetään tärkeänä välineenä kerätä ja tarkastella tietoa yhteiskunnan ilmiöistä sekä mm. ihmisten mielipisteistä, asenteista ja arvoista. (Vehkalahti, 2008, s.11.)

Ongelman tai tutkimuskysymyksen muuntaa mitattavaan muotoon kutsutaan operationalisoinniksi (Vehkalahti, 2008, s.18). Tässä tutkimuksessa kyselylomake toimii tutkimuskysymykseen liittyvien asioiden mittausvälineenä. Mitattavat aiheet ovat abstrakteja ja niitä mittaamaan suunniteltiin konkreettisia kysymyksiä (emt.).

Kyselytutkimukselle sekä kyselylomakkeen kysymyksille eli mittareille on ominaista epävarmuus ja tutkimuksen onnistumisen kannalta on keskityttävä moneen tekijään. Pitää tarkastella edustivatko kyselyyn osallistuneet perusjoukkoa, saatiinko heiltä tarpeeksi vastauksia ja vastasivatko he riittävän tarkasti. Lisäksi on hyvä pohtia mittasivatko kysymykset tutkittavia asioita ja toimivatko mittarit luotettavasti. (Vehkalahti, 2008, s.11.)



Kuvio 3.1. Kyselytutkimuksen prosessi

Lähde: *Laatua tilastoissa*. Tilastokeskus (2007, s.29).

Kyselytutkimuksen eli *survey*n vaiheet koostuvat suunnitteluvaiheesta, operationalisoinnista sekä arviointivaiheesta (kuvio 3.1). Jokainen vaihe on tärkeä ja jokaisessa vaiheessa on monta tekijää, jotka pitää ottaa huomioon mahdollisimman hyvän kyselytutkimuksen aikaansaamiseksi.

Kysely tai tiedustelu voidaan lisäksi toteuttaa monenlaisin välinein, kuten postitse,

puhelimitse, käyntihaastatteluun tai sähköisen viestimen kuten internetin avulla (Laaksonen, 2009). Menetelmät eroavat ominaisuuksiltaan toisistaan kustannusten ja huomioon otettavien erityispiirteiden suhteen.

Tämän tutkimuksen suunnittelu sekä toteuttamistapa syntyi yhteistyössä työnohjaajan Kimmo Vehkalahden sekä AYY:n edustajan Hanna Saulin kanssa. Ideoinnin seurauksena päätimme lähestyä tutkittavaa aihetta vertaamalla lukukausimaksukokeiluun valituissa maisteriohjelmissa vuosina 2010 ja 2011 aloittaneita opiskelijoita toisiinsa. Tarkoituksena oli siis verrata kahta mahdollisimman samanlaista joukkoa, sillä erotuksella, että vuonna 2011 aloittaneista opiskelijoista osa maksaisi lukukausimaksuja. Lisäksi kaikki vuonna 2011 hakeneet EU- ja ETA-maiden ulkopuoliset opiskelijat tiesivät hakiessaan maksukokeilusta ja hakivat näin tiedostaen maksujen mahdollisuuden stipendijärjestelmästä huolimatta. Tutkimus oli helppoa ja selkeintä toteuttaa kyselytutkimuksena verkossa, jolloin kustannukset olisivat mahdollisimman matalat ja lisäksi hyvin sensitiivisiin kysymyksiin opiskelijan sosioekonomisesta taustasta saataisiin mahdollisesti parempia vastauksia kuin haastattelijan läsnäollessa.

3.1 Kyselylomakkeen suunnittelu

Tutkimuksen kannalta kyselylomake on ratkaisevassa asemassa ja siten koko tutkimuksen onnistuminen riippuu hyvin paljon lomakkeesta. Kysymyslomaketta tehdessä on tärkeää pitää mielessä, että kysytään sisällöllisesti oikeita kysymyksiä tilastollisesti mielekkäällä tavalla. Lomakkeen suunnitteluun ja mittausvaiheeseen kannattaa siis panostaa todella, sillä tehtyjä virheitä ei voida millään analyysimenetelmällä korjata tai oikaista. (Vehkalahti, 2008, s.20.)

AYY:n esittämään tutkimusongelmaan perustuva kyselylomake ideoitiin yhdessä AYY:n edustajan Hanna Saulin kanssa. Ideointivaiheeseen kuului kysymyslomakkeella tutkittavien ulottuvuuksien hahmottaminen, kohderyhmään perehtyminen ja esimerkkikysymysten laatiminen tässä järjestyksessä. Pohdimme myös AYY:n kannalta sopivaa lähestymistapaa kysymysten asetteluun, jotta tutkimuksen henki ja ote säilyisi sopivan opiskelijäläheisenä.

Kyselytutkimus on usein määrällistä tutkimusta eli kyselyaineistot koostuvat pääosin mitatuista luvuista, vaikka kysymykset esitetään sanallisesti (Vehkalahti, 2008, s.13). Tässä tutkimuksessa asenteita mitattiin erilaisilla väittämillä, joihin vastaaja pystyi ottamaan kantaa viisiportaisella lineaarisella asteikolla samaa mieltä – eri mieltä (*Totally agree – Totally disagree*). Vastaukset koodattiin aineistoon kuitenkin numeroilla 1 – 5.

Kyselytutkimuksissa sanallinen tieto on usein täydentävää tietoa tai vastauksia kysymyksiin, joihin vastaaminen numeerisesti olisi epäkäytännöllistä (Vehkalahti, 2008, s.13). Tässä tutkimuksessa jokaisen väittämäpatteriston jälkeen on ollut tilaa ilmaista itseään sanallisesti, jos väittämät eivät sopineet vastaajalle, niissä oli jotain hankalaa tai selvennettävää, tai vastaajalla oli muuten kommentteja asiasta.

Seuraavissa osioissa käyn läpi tämän kyselylomakkeen kannalta tärkeitä asioita ja keskityn erityisesti internet-lomakkeella tehtävään kyselyyn ja sen erityispiirteisiin.

3.1.1 Mittausmalli

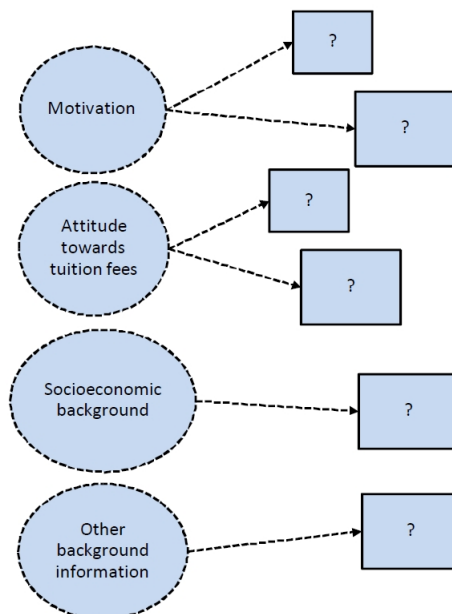
Mittausmallin rakentaminen on hyvä tapa aloittaa kysymysten suunnittelu. Mittausmalli on esimerkki tilastollisesta mallista ja se täsmentää mitä mitataan ja miten. Mallia

rakentaessa tutkija joutuu pohtimaan tutkimusongelmaa syvällisemmin ja tutkijalle selkiytyy paremmin mitä hän haluaa oikeasti mitata. Mittausmallissa lähdetään liikkeelle tutkimusongelmasta tai -kysymyksestä. Mietitään mitkä ovat ilmiön tai ongelman ulottuvuudet ja millaisilla kysymyksillä ulottuvuuksia voisi mitata. Ulottuvuuksia lähdetään miettimään siis ensisijaisesti tutkimuksen sisällölliseltä kannalta. Osioita eli kysymyksiä on oletusarvoisesti useampia kuin ulottuvuuksia ja ne voivat liittyä useaan mietittyyn ulottuvuuteen. (Vehkalahti, 2008, s.20.)

Seuraavaksi kerron tämän tutkimuksen kysymyslomakkeen muodostumisesta ja rakenteesta mittausmallin avulla. Huomautan, että en ole kääntänyt mittausmallin ulottuvuuksia tai kysymyksiä suomenkielelle vaan esitän ne tässä työssä alkuperäiskielellään englanniksi. Kysymyksiä käännettäessä on aina riskinä kysymysten sisällöllinen muutos, mikä aiheuttaa vääristymiä ymmärryksessä sekä siis vastauksissa. Kääntäminen tulisi suorittaa aina monen asiantuntijan toimesta, niin että käännös varmasti vastaa sanamuodoltaan sekä sisällöltään alkuperäistä (Harkness, Pennell & SchouaGlusberg, 2004, s.453–473). Kysymyksiä ei tarvita tässä tutkimuksessa kuin englanniksi, joten kääntäminen olisi ollut resurssien tuhlaamista.

Ulottuvuudet

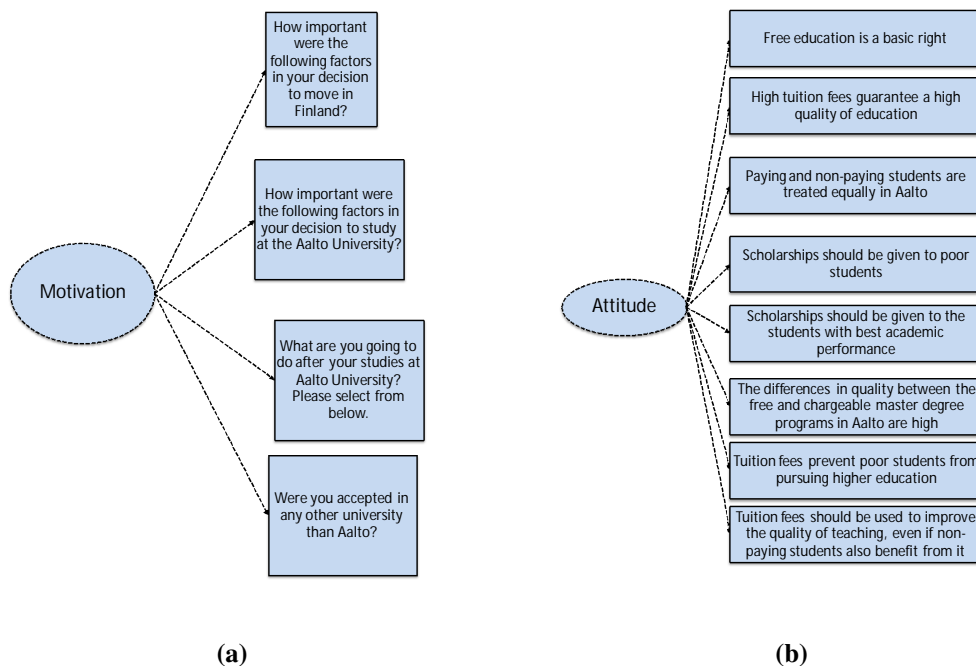
Mittausmalli lähtee ulottuvuuksista eli asioista tai ilmiöistä mitä halutaan tutkia. Tutkimuskysymystä pohdittuani päädyin kolmeen ulottuvuuteen, jotka olivat motivaatio tulla Suomeen, asenne ja sosioekonominen tausta. Analyysien tekemiseksi ja laadun varmistamiseksi tarvitaan lisäksi taustatietoja kuten ikä ja sukupuoli. Ulottuvuudet on esitetty kuviossa 3.2 katkoviivoilla merkittyjen ympyröiden sisällä. Jokaisesta ympyrästä lähtee viiva laatikoilla merkittyihin mittareihin eli kysymyksiin, jotka ovat vielä tässä vaiheessa tuntemattomia.



Kuvio 3.2. Ulottuvuudet

Kysymykset eli osiot

Kysymysten laadinnassa lähdin edellä mainituista ulottuvuuksista, jotka olivat monimutkaisia sekä hyvin erilaisia luonteiltaan. Erilaisten väittämien laadinta tuntui sopivimmalta vaihtoehdolta kaikkia ulottuvuuksia ajatellen. Lisäksi analysointimenetelmiä pohdittuani väittämät olivat paras vaihtoehto tutkimuskysymysten operationalisoinniksi ja näin muuttujien asteikoista saatiin yhtenäiset koko lomakkeeseen. Jokaisesta ulottuvuudesta tuli keksiä monta kysymystä tai väittämää, sillä näin subjektiivisiin kysymyksiin liittyy suurempi epävarmuus (Laaksonen, 2009). Seppo Laaksonen listaa luentomonisteesaan (Laaksonen, 2009, s.15) seitsemän peruskysymystä, joihin on hyödyllistä vastata lomaketta laadittaessa. Näillä ohjaavilla kysymyksillä pyritään siihen, että kysymysten suunnittelija miettisi kysymysten relevanttiutta, tarkoitusta, hyvyttä sekä toimivuutta. Kysymyksiä tehdessä ja muotoillessa koetin käyttää näitä kysymyksiä ohjenuorana.

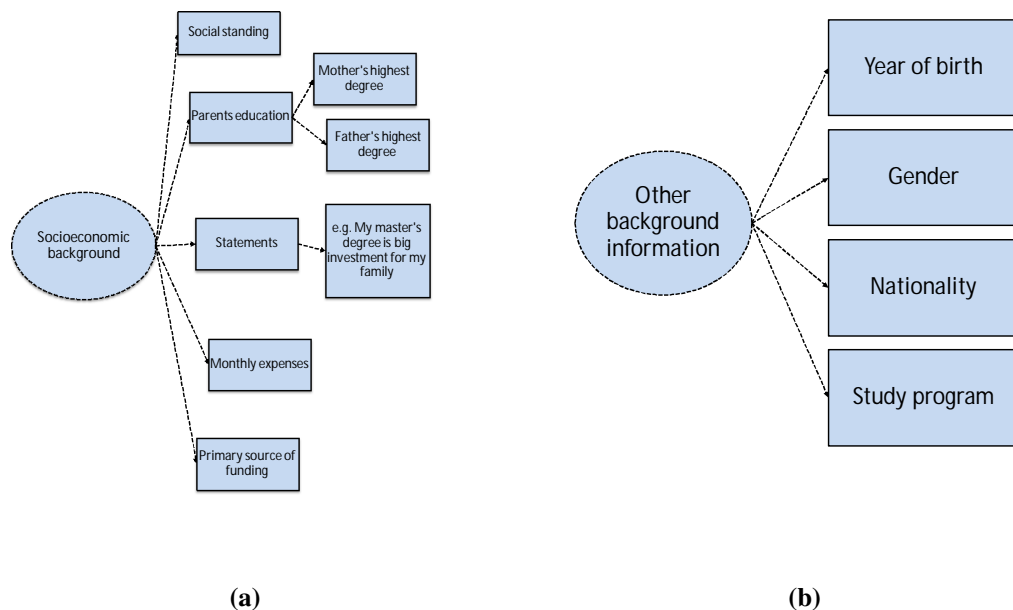


Kuvio 3.3

Kuvassa 3.3a on esitettyä ulottuvuus nimeltään motivaatio (*Motivation*) eli tekijät, jotka houkuttelivat opiskelijan Suomeen ja Aalto-yliopistoon. Kuvion neliöissä on ulottuvuutta kuvaamaan suunnitellut todelliset kysymykset eli muuttujat. Suomeen ja Aalto-yliopistoon liittyviin väitteisiin sain materiaalia Johanna Kärjen selvityksestä¹. Kaikki väittämät on mitattu viisiportaisella asteikolla välillä *Not important* – *Very important* ja lisäksi vaihtoehtona oli vastata *Don't know* eli en tiedä. Väitteitä ei näy kuvasta 3.3a vaan niitä voi tarkastella kyselylomakeen (liite D) osasta 2. Ulottuvuuteen kehitettiin kaksi lisäkysymystä koskien tulevaisuuden suunnitelmia sekä Aalto-yliopistoon hakeutumista.

Asenteita (*Attitude*, kuva 3.3b, liite D lomakkeen sivu 1) mittaamaan keksimme yhdessä Hanna Saulin kanssa yhteensä kahdeksan väittämää. Väittämät vaihtelevat myös

¹ If I had to pay I would require value for my money, 2005



Kuvio 3.4

viisiportaisella asteikolla, jossa vastaajalle on näytetty ääripäät *Totally agree – Totally disagree* sekä *I don't know* -vaihtoehto.

Sosioekonomista taustaa mittaaviin kysymyksiin hain mallia Tilastokeskuksen², European Social Surveyn (ESS)³ sekä Suomessa tehdystä International Student Barometer (ISB) -tutkimuksesta⁴. Kysymyksiksi valikoituivat vanhempien koulutustaso ja pääasiallinen toimi sekä subjektiivinen kuva perheen sosiaalisesta asemasta (kuvio 3.4a).

Taustatekijöiksi valikoituivat sukupuolen ja iän lisäksi Suomeen muuttovuosi, kansallisuus, koulutusohjelman nimi sekä aloitusvuosi Aalto-yliopistossa (kuvio 3.4b).

Kysymyspatteristot ovat nähtävissä kokonaisuudessaan liitteestä D.

3.1.2 Kehikkoperusjoukko ja sen asettamat vaatimukset

AYY halusi tutkia maksukokeilun aiheuttamia muutoksia, joten tarkasteluun oli otettava maksukokeiluun osallistuvien maisteriohjelmien opiskelijat. Maksavien ja ilmaiseksi opiskelevien vertailtavuus suunniteltiin ottaa huomioon siten että kysely kohdennettaisiin myös 2010 aloittaneille opiskelijoille. Lisäksi kiinnostuksen kohteena olivat maksukokeiluun osallistuvien maisteriohjelmien EU- ja ETA-maiden ulkopuoliset opiskelijat, joten kehikoksi muodostui 2010 ja 2011 aloittaneet lukukausimaksukokeiluun osallistuvien maisteriohjelmien EU- ja ETA-maiden ulkopuoliset opiskelijat.

Kehikkoa kuitenkin laajennettiin, kun tieto maksavien opiskelijoiden hyvin pienestä määrästä tuli selville ja tutkimuksen muututtua pilottimaiseksi koskemaan kaikkia englanninkielisiä maisteriohjelmien ulkomaalaisia opiskelijoita Aalto-yliopistossa.

²http://www.stat.fi/meta/kas/sosioekon_asema.html

³<http://www.europeansocialsurvey.org/>

⁴<http://www.i-graduate.org/services/international-student-barometer-and-student-barometer/>

AYY on opiskelijan etua ajava järjestö, siispä tutkimuksesta haluttiin tehdä opiskelijoiden näköinen. Turha kankeus pyrittiin minimoimaan ja kysymysten asettelussa pyrittiin säilyttämään rentous sekä keskusteleva ote. Pro gradu -työni ohjaajan Vehkalahden ehdotuksesta kysely ei ala usein vastaavissa kyselyissä nähtävillä taustakysymyksillä, vaan lomake alkaa mielipidekysymyksillä. Näin vältetään kuulusteleva ote ja vastaajan mielenkiinto aiheeseen herätetään. Taustakysymyksiä, kuten ikää ja sukupuolta, kysytään lomakkeen loppupuolella.

Lomakkeen kieleksi valittiin luonnollisesti englanti, sillä opiskelijat ovat tulleet opiskelemaan englanninkielisiin maisteriohjelmiin. Opiskelijoilla on siis todennäköisesti hyvät taidot englanninkielessä. Lomakkeen kysymykset haluttiin pitää kuitenkin mahdollisimman yksinkertaisina, jotta kaikki ne ymmärtäisivät, eikä turhia, esimerkiksi kulttuurieroista johtuvia, väärinkäsityksiä syntyisi. Lisäksi liian hienot sanamuodot tai vaikeat kielelliset rakenteet olisivat osaltaan vieneet pois opiskelijäläheistä tunnelmaa lomakkeesta.

Aalto-yliopiston tiloissa on tarjolla opiskelijoille monia tietokonesaleja ja jokainen saa niihin käyttäjätunnukset. Lisäksi langaton internet on käytössä koko kampusalueella. Jokaisella opiskelijalla pitäisi siis olla mahdollisuus tietokoneeseen ja internet-yhteyden käyttöön joten kaikilla on teoriassa mahdollisuus osallistua myös verkossa toteutettavaan tutkimukseen.

Lisäksi on todennäköistä, että monella opiskelijalla on myös oma tietokone sekä internet-yhteys käytettävissään myös korkeakoulun ulkopuolella. Opiskelijat ovat tottuneita tietokoneiden käyttäjiä ja asioivat tottuneesti internetissä. He etsivät tietoa, kommunikoivat ystävien kanssa sekä viettävät aikaa mm. blogeja lukien. (Suomen virallinen tilasto (SVT): Tieto- ja viestintätekniikan käyttö, 2011)

Näillä oletuksilla päätin toteuttaa kyselyn internet-pohjaisena verkkokyselynä. Verkossa tehtävästä kyselystä kerron tarkemmin seuraavassa osiossa.

3.2 Verkkolomake tiedonkeruuvälineenä

Tässä osiossa kerron verkkolomakkeen ominaisuuksista sekä seikoista, jotka tulisi ottaa huomioon verkkolomaketta laatiessa. Lisäksi kerron lyhyesti käyttämästäni e-lomakepalvelusta ja tämän tutkimuksen kannalta tärkeistä seikoista.

Verkkolomakkeet tarjoavat paljon uusia mahdollisuuksia kyselylomakkeiden maailmassa. Ne eroavat hyvin paljon perinteisistä postitettavista paperilomakkeista sekä puhelinhaastatteluista interaktiivisen luonteensa ja itseohjautuvuutensa vuoksi. Internet on myös itsessään oma erillinen ympäristönsä eikä sitä voida verrata mihinkään muuhun ympäristöön. (Couper, 2008, s.28.)

Yksi suurimmista internet-lomakkeen hyödyistä, niin ajallisesti kuin taloudellisestikin, on se, että vastaaja täyttää itse lomakkeen sähköisesti. Aineistonkeruu ja -tallennus helpottuvat huomattavasti kun vastaaja itse tallentaa vastauksensa suoraan palveluun ja kaikki vastaukset ovat kyselyn päätyttyä valmiina samassa internet-tietokannassa reaaliaikaisesti. (emt.)

Tämän lisäksi internet-lomakkeen kustannuksia alentaa sen virtuaalinen toistettavuus. Kun lomake on kerran laadittu, sen levittämiseen ei tarvita tulostusta tai perinteistä postitusta, joten kyselyn teettämisestä ei kerry lisäkustannuksia. Data on valmiina reaaliaikaisesti eikä sen tallennus tai keräys ole sidottuna haastattelijoihin. Lisäksi erilaiset

tarkistusalgoritmit vähentävät vastaajien inhimillisiä virheitä. (emt.)

Lomaketta on myös mahdollisuus muuttaa reaaliaikaisesti, jos esimerkiksi jokin osa lomakkeesta ei ole toiminut tai jossain huomataan vakava kirjoitusvirhe.

Verkkolomakkeen heikkoutena on se ettei vastaajaa ole ohjaamassa ketään, jolloin ongelman tullessa vastaaja jättää kyselyn kesken tai vastaa epätotuudenmukaisesti. Tätä kutsutaan vastauskadoksi ja se on verkkokyselyn suurimpia ongelmia. Verkkokyselyn ongelmana on ollut myös perusjoukon peittävyysongelmat, sillä kaikilla ei ole pääsyä internetiin (Börsch-Supan, Elsner, Fassbender, Kiefer, McFadden & Winter, 2004). Tutkimusten mukaan internetiin pääsyn todennäköisyys riippuu iästä, koulutustasosta ja sukupuolesta niin, että nuorilla koulutetuilla miehillä on korkein todennäköisyys käyttää internetiä ja ovat näin ylliedustettuina esimerkiksi koko kansaa koskevissa verkkokyselyissä (emt.). Tässä tutkimuksessa tämän peittävyysongelman oletetaan olevan olematon, sillä opiskelijat ovat ahkeria tietokoneen ja internetin käyttäjiä.

Verkkolomakkeen tulisi olla hyvin käyttäjäkeskeinen eli kysymysten ja tehtävien tulisi olla helppoja sekä miellyttäviä vastaajalle, eikä vastaajan tulisi turhaan kiinnittää huomiota kyselyn prosesseihin. Lisäksi aiheen tärkeys ja pätevyys ovat tärkeitä kannustimia vastaajan näkökulmasta. (Couper, 2008.)

Jos kyselylomaketta lähdetään rakentamaan alusta lähtien itse, tarvitaan ohjelmointitaitoja, jotta kysymykset saataisiin tietokonekielelle ja vastauksia varten pystytettäisiin luomaan elektroninen tietovarasto (Couper, 2008). On myös olemassa valmiita internetissä toimivia palveluita, kuten Helsingin yliopiston käyttämä e-lomake-palvelu, jossa kaikki nämä asiat on mietitty valmiiksi. Lomakkeen pystyy kokoamaan helposti erilaisista osista graafisen ja web-pohjaisen käyttöliittymän avulla.

3.2.1 Huomioon otettavaa verkkolomaketta laatiessa

Verkkolomakkeen ulkoasua ja toimivuutta ei voida kontrolloida loputtomiin, sillä eri tietokoneiden, selainten ja asetusten yhdistelmiä on tuhansia, ja kaikissa näissä lomake näyttää hiukan tai paljonkin erilaiselta kuin muissa kombinaatioissa (Couper, 2008). Kaikissa tietokoneissa on omat valmistajalleen ominaiset sekä käyttäjänsä laatimat asetukset, jotka kontrolloivat mm. sivujen esitystä, hiiren nopeutta ja näytön kirkkautta. Vastaajan internet-yhteyden nopeus voi vaihdella modeeminopeuksista kuitulaajakais-toihin, jolloin mahdollisten sivujen lataus tai kuvien esitys ei näy samanlaatuisena kaikille. Ei ole myöskään olemassa yhtä ja ainoata internet-selainta, ja pelkästään sama selain eri käyttöjärjestelmissä toimii erilailla ja näyttää erilaiselta. Se, että lomake näyttää edes kohtuullisen samanlaisena kaikille vastaajille, on tärkein tavoite verkkolomaketta laatiessa (Baker, Crawford & Swinehart, 2004, s.361–384). Lomaketta tulisikin siis testata mahdollisuuksien mukaan eri kombinaatioilla ja ottaa mahdolliset eriävyydet huomioon (emt.). Lomaketta testattiin sekä Windows että Mac -käyttöjärjestelmissä eri selaimilla. Kaikissa kombinaatioissa lomake näytti suhteellisen samanlaiselta.

Vastaajia voi arveluttaa internetin **anonymiteetti** ja se että kuka vaan voi esiintyä tutkijana. Roska- ja huijausposteja tulee nykyään päivittäin, eikä omia tietoja maalaisjärjen mukaan kannata luovuttaa mihin tahansa. Tuleviin viesteihin ja kyselyihin tutustutaan siis varauksella. Lisäksi lisääntyneen roskapostin vuoksi monella on käytössä hyvin vahvat roskapostinsuodattimet ja kyselystä kertova viesti voi hyvin ohjautua tarkoituksettomasti roskapostikansioon ja aiheuttaa näin vastauskatoa tutkimuksessa. Toisaalta anonymiteetti heikentää sosiaalisesti hyväksyttäviä vastauksia ja vastaajat ovat keski-

määrin rehellisempiä vastauksissaan. (Dillman, 2009, s.195, 232–233.)

Verkkolomake on **itse ohjautuva** (*self administered*), sillä kukaan ei ole neuvomassa vastaajaa lomakkeen kanssa. Näin vältetään haastattelijavaikutukselta, mutta haastattelijan puuttuessa puuttuu myös tuki vaikeuksia kohdatessa. Lomakkeen tulee siis olla tarpeeksi selkeä, jotta vastaaja pystyy sen ilman avustusta täyttämään eikä turhaudu. Lisäksi mahdollisten virheiden varalta tulisi olla järjestelmä, joka ohjaa vastaajaa ratkaisemaan ongelmat. Itseohjautuvuus takaa myös sen, että vastaajalla on mahdollisuus edetä haluamallaan vauhdilla kuten paperilomaketta täyttäessään. Internet-yhteys tarjoaa kuitenkin laajan kirjon eri häiriötekijöitä ja lomakkeeseen keskittyminen ja sen loppuun täyttäminen voi olla vaikeaa. (Couper, 2008, s.28.)

Verkkolomake on **interaktiivinen** ja tätä ominaisuutta tulisi käyttää hyödyksi myös vastaajaa motivoidessa osioiden välillä (Couper, 2008). Kannustavat tekstit tai kuvat saattavat vähentää puuttuvia vastauksia. Tässä kyselyssä lomake haluttiin pitää mahdollisimman yksinkertaisen, jottei vastaajan keskittyminen menisi epäoleellisiin asioihin ja kuvat tms. jätettiin kokonaan pois. Myös monimutkaisilla tarkistuskomennoilla voidaan selvittää miksi joku on vastannut kaikkiin ”en tiedä”, keskeyttänyt ja muutenkin tehdä laadunvalvontaa jo vastaajan täyttäessä lomaketta.

Sivujen ryhmittely tai miten se tehdään, on yksi tärkeimmistä päätöksistä verkkolomaketta tehdessä. Jos kaikki kysymyksen esitetään yhdellä internet-sivulla, lomake muistuttaa paperista lomaketta. Yhden sivun mallin hyviä puolia on mm. se että vastaaja näkee sivua vierittämällä kyselyn kokonaisuutena, joka auttaa hahmottamaan vastaamiseen kuluva työtaakkaa. Yhden sivun mallissa vastaaja kuitenkin ohittaa huomaamattaan helposti kysymyksiä tai kokonaisia osioita, varsinkin, jos kysely on kovin pitkä. Lisäksi itselleen epärelevanttien kysymysten ohittaminen voi olla vaikeaa kuten paperisessakin lomakkeessa ilman hyviä ohjeita. Vastauksista ei myöskään tallennu ollenkaan dataa, jos vastaaja ei tee kyselyä loppuun ja selainasetukset estävät evästeiden käytön. Yksisivuisuus estää interaktiiviset toiminnot kuten edellisten kysymysten vastauksista muodostetut luvut tai spesifioidut kysymykset, sillä sivu pysyy muuttumattomana. (Couper, Peytchev, McCabe & Crawford, 2006, s.604.)

Kun lomakkeen kaikki kysymykset esitetään eri sivuilla, mahdollistetaan monimutkaisetkin ohitusmallit ja vastaajan vastuu ja taakka kevenee. Lisäksi vastaajan huomio ja keskittyminen saadaan tarkennettua aina käsillä olevaan kysymykseen. Jos jokainen kysymys on eri sivulla, vastaajalla kuluu enemmän aikaa ”seuraava sivu” nappien painamiseen ja seuraavan sivun latautumista odotellessa. Eri sivut kysyvät vastaajalta myös lisätyötä, jos henkilön tulisi muistaa vastauksiaan edellisissä kysymyksissä. (emt.)

Kahta edellistä yhdistelemällä, eli kokoamalla toisiinsa liittyvät kysymykset eri sivuille, säästetään vastaajan aikaa, mutta rakenne pysyy selkeänä ja lomakkeeseen voidaan asettaa ehdollisia kysymyksiä, jotka mahdollistavat epärelevanttien kysymysten ohittamisen. Kun kysymyksiä kootaan kokonaisuuksiksi sivuille, on muistettava pitää yhdellä sivulla vain oikeasti toisiinsa liittyvät kysymykset. Vastaajilla on taipumus yrittää keinoitekoisesti yhdistellä asioita ja näin vastaaja voi hämmentyä ja/tai vastaukset voivat vääristyä. (emt.)

3.2.2 E-lomakepalvelu

Helsingin yliopisto tarjoaa opiskelijoidensa ja työntekijöidensä käyttöön e-lomake palvelun, jota päätin käyttää työssäni. Palvelussa voi tehdä internet-pohjaisen verkkolomak-

keen käyttäen valmiita kysymyspohjia monivalinnoista ja hierarkisista valikoista avoimiin kenttiin. Lomakkeen ulkoasua voi myös muokata omannäköiseksi joko valmiista vaihtoehtoista valitsemalla tai kehittämällä sopivan koodin. Päädyin käyttämään valmista ulkoasua tässä kyselyssä.

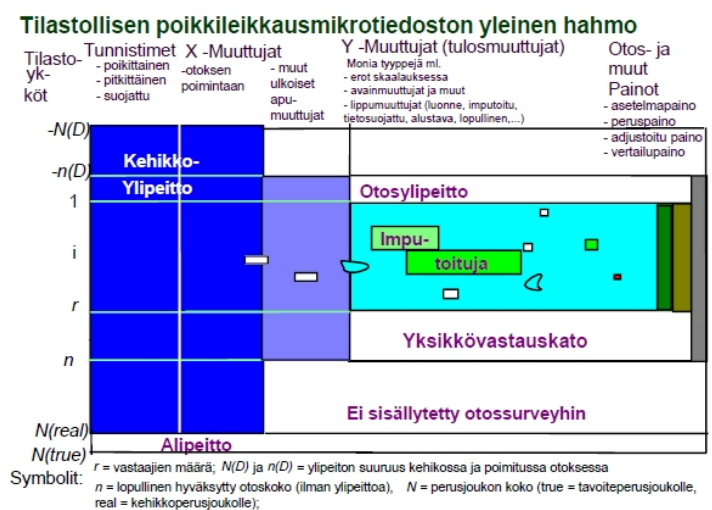
E-lomakepalvelun käyttöä puolsivat sen kysymysvaihtoehtojen ja toimintojen monipuolisuus. Palvelu tuntui myös turvalliselta käyttää, sillä Helsingin yliopisto on ostanut käyttöoikeudet siihen, ja palvelu on tullut tutuksi useasta yhteydestä yliopiston eri tahojen käyttäessä sitä moniin tarkoituksiin. Kyselylomakkeeseen oli myös mahdollista liittää salasana, jolla estetään kohderyhmään kuulumattomien vastaukset ja näin turha ylipeitto. Palvelun avulla pystytään myös lähettämään muistutusviesti vain niille, jotka vielä eivät ole vastanneet. Salasanallisuus lisää tutkimuksen tulosten luotettavuutta kun tiedetään, että vastaajat kuuluvat kohderyhmään. Se lisää myös luottamusta myös vastaajissa, sillä näin syntyy vaikutelma, etteivät tiedot ole missä tahansa esillä.

Kun kyselyyn osallistuva vastaa e-lomakkeeseen, tallentuvat tiedot reaaliaikaisesti verkkopalvelun palvelimelle, mistä ne ovat ladattavissa Excel-tai SPSS-tiedostoksi datan käsittelyä varten.

Lomaketta tehdessä vastaan tuli muutama e-lomakepalvelun ominaisuus, jotka jossain määrin rajoittivat työtäni. Tein lomakkeen englannin kielellä, ja huomasin, ettei kaikkia toimintoja löytynyt englanniksi. Huomautin asioista lomakepalvelun ylläpitoon ja kyseiset epäkohdat korjattiin onneksi nopealla aikataululla. Myös lomakepalvelun kysymysvaihtoehtojen (avoin, radionappi yms.) sekä sivun sommittelumahdollisuuksien jähkyys ilmeni työn edetessä. Tästä syystä jouduin tekemään joidenkin kysymysten kohdalla kompromisseja selkeyden, käytettävyyden ja ymmärrettävyyden välillä.

3.3 Vastauksadon ennaltaehkäisy

Vastauksato syntyy siitä, että kohderyhmän edustaja ei vastaa kyselyyn ollenkaan tai vastaa siihen puutteellisesti. Ensimmäistä kutsutaan yksikkökadoksi ja toista erävastaukskadoksi (Laaksonen, 2009, s.112). Niiden vaikutukset koko aineistoon ovat havaittavissa kuviosta 3.5.



Kuvio 3.5. Vastauksato aineistossa. Lähde: Laaksonen (2009, s. 45)

Yksikkövastauskato on yleistä varsinkin verkkokyselyissä, sillä kyselyitä saapuu sähköpostiin jatkuvasti. Internet-kyselyitä on liikkeellä paljon, sillä niitä voi levittää helposti napinpainalluksella ja vielä ilman lisäkustannuksia. Vastaanottajat ovat uupuneita kyselyiden määrään ja vastaavat vain kaikkein mielenkiintoisimpiin ja motivoivimpiin kyselyihin. On siis tärkeätä, että vastaaja tuntee olevansa avainasemassa tutkimuksen kannalta, ja että kyselyyn vastaamista on motivoitu jollain keinolla. (Lynn, 2008, s.41.)

Toinen yksikkövastauskatoa aiheuttava tekijä internet-kyselyissä on vanhentuneet ja/tai väärät sähköpostiosoitteet, jotka aiheuttavat ali- ja ylipiteittoa (ks. kuvio 3.5). Tässä tapauksessa lähetetty kutsu ei koskaan edes saavuta kohderyhmään kuuluvaa ja syntyy alipeittoa tai viesti saapuu kohderyhmään kuulumattomalle ja syntyy ylipiteittoa. Osoitelista tulisinkin olla mahdollisimman ajan tasalla sekä tarkka. (Lynn, 2008, s.41.)

Erävastauskadon syiksi on eritelty kolme vaihtoehtoa. Joko vastaaja ei ymmärrä mitä kysymyksellä haetaan, hän ei tosiaan tiedä (tarpeeksi hyvin) kysyttyä asiaa tai hänellä ei ole halua tai motivaatiota luovuttaa kysyttyä tietoa. Niinpä jokaisen kysymyksen tulisi vähintään olla ymmärrettävä ja vastaajalle olisi tarpeen tarjota joustavia vastausvaihtoehtoja (Groves, Fowler, Couper & Lepkowski, 2004, s.188.)

Kysymyksiin liittyvistä tekijöistä kerron lisää kohdassa 3.3.2

3.3.1 Kyselytutkimukseen liittyvät yleiset tekijät

Yksi hyvin kustannustehokas keino lisätä vastausprosenttia on **etukäteistiedonanto** tutkimuksesta (Couper, 2008, s. 306). Kohderyhmään kuuluva saadessaan kutsun todennäköisemmin avaa myös sen, sillä tutkimus on jo tuttu vastaajalle. Tämän tiedon perusteella sovimme, että AYY:n Hanna Sauli kertoo pitämässään AYY:n infotilaisuuksissa kohderyhmän opiskelijoille tulevasta tutkimuksesta. Tämä etukäteistieto ei varmasti tavoittanut kaikkia kohderyhmäläisiä, mutta uskoimme siitä olevan kuitenkin hyötyä vastausaktiivisuuden kannalta.

Saatekirje toimii tutkimuksen julkisivuna, jonka tarkoituksena on houkutella ja motivoida vastaanottaja osallistumaan kyselyyn. Saatekirje on siis avainasemassa vastaajien rekrytoinnissa. Tässä tutkimuksessa saatekirjeenä toimii sähköpostiviesti, jonka ominaisuudet eroavat tavallisesta kirjeestä (Couper, 2008, s.308–310).

Ajankohtainen luettelo yhteystiedoista on ratkaisevassa asemassa, jotta alipeittoa välttyttäisiin. Toiseksi on varmistettava asiallisilla sanavalinnoilla ja lähettäjäosoitteella, ettei viesti jää roskapostinsuodatuksen, vaan saapuu vastaanottajan postilaatikkoon.

Kuten kuvasta (3.6) nähdään, tärkeimmät kyselyn kutsun ensivaikutelman tekijät ovat lähettäjä sekä aihe. Jos nämä eivät ole kunnossa, kohderyhmän edustajaa ei saada edes aukaisemaan viestiä, ja kysely menettää heti vastaajan. Lähettäjän tulisi edustaa kyselyn kannalta relevanttia tahoa ja osoitteen tulisi oikeasti olla olemassa (Couper, 2008, s.321). Jotkut voivat lähettää suoraan palautetta tutkijalle ja tällöin on sopivaa, että vastaus menee perille ja oikeaan osoitteeseen. Päädyimme käyttämään tutkimuksen tekijän helsinki.fi -loppuista sähköpostiosoitetta. Näin kysely edusti selvästi ulkopuolista tahoa Aalto-yliopistoon nähden ja mahdollisille vastauksille sekä kommentteille saatiin oikea osoite.

Aiherivin tulisi olla mielenkiintoinen ja kutsuva, ja sitä tulisi varioida, jotta se vetoaisi mahdollisimman moneen (Couper, 2008, s.310). Tässä tutkimuksessa otsikkoa varioitiin eri muistutuskierröksillä ja otsikot ovat nähtävissä liitteestä C.

Sähköpostilistat antavat sellaisen kuvan, että vastaanottajia on tuhansia ja juuri tä-



Kuvio 3.6. Sähköpostikutsun ominaisuuksia. Vapaasti suomentaen (Couper, 2008, s. 306)

män vastaajan vastauksella tuskin on merkitystä. Koko vastaanottajalista vastaanottajakentässä taas rikkoo yksityisyyttä ja viestii myös massapostituksesta. Vastaanottajakentässä tulisi siis olla ainoastaan kyseisen vastaanottajan tiedot (Couper, 2008, s.311). Tätä yksilöintiä eli personointia käytetään mm. markkinoinnin keinona markkinointiviestien tehostamiseksi. Tekniikalla on Dillmanin (1991, s.225) mukaan pieni parantava vaikutus vastausprosenttiin.

Viimeisenä tulisi varmistaa, että kutsussa oleva linkki tai osoite kyselyyn on varmasti oikea ja toimii (Couper, 2008, s.322). Tarkistin asian lähettämällä kutsun ensin itselleni.

Tämän tutkimuksen saatekirje muotoiltiin yhdessä AYY:n Hanna Saulin kanssa pyrkien noudattamaan mainittuja ohjeita. Valmis saatekirje löytyy liitteistä (Liite C). Otsikkovariaatiot on merkitty liitteeseen nimillä 'Vaihtoehto 1' jne.

Muistuttelu on tehokkain vastausprosenttia lisäävä tekijä. Liian useata muistutusta ei kuitenkaan tule lähettää, sillä vastaanottajat voivat pahastua jatkuvasta viestittelystä. Tätä ilmiötä kutsutaan kyselyväsymykseksi tai -ärtymykseksi. Lisäksi muistutusviestien teho laskee jyrkästi ensimmäisen ja toisen muistutusviestin jälkeen. Tässäkin tilanteessa tulisi käyttää mahdollisuuksien mukaan personointia niille, jotka eivät vielä ole vastanneet ollenkaan ja niille, jotka ovat jättäneet kyselyn kesken. (Couper, 2008, s.340.)

Muistuttelun vaikutuksia voi arvioida kuviosta 3.7 sivulta 37. Väsymysilmiö oli havaittavissa jo toisen muistutusviestin kohdalla, sillä sain henkilökohtaisia sähköposti-viestejä, joissa kieltäydyttiin vastaamasta kyselyyn.

3.3.2 Kysymysten ominaisuudet

Lomakkeen tunnelma haluttiin pitää opiskelijamaisena valitsemalla mahdollisimman yksinkertaisia kysymyksiä sekä sanoja. Yksinkertaisuudella oli myös toinen tarkoitus. Kysymysten helppoudella ja lyhyydellä haluttiin välttää vastausväsymys sekä kielelliset sekaannukset. Kyselyt ovat usein liian pitkiä ja vieraat sanat tai monimutkaiset rakenteet vievät huomion pois oleellisesta eli kyselystä (mm. Dillman, 2000). Vaikka opiskelijat ovat tulleet opiskelemaan englanniksi, on englanti harvan äidinkieli. Monimutkaiset sana- ja lauserakenteet olisivat lisänneet sekaannuksia sekä tulkintavirheitä.

Jotta vastaaja ymmärtäisi kysymykset oikein ja osaisi tuottaa totuudenmukaisia vastauksia, tulisi jokaisen kysymyksen sekä vastausvaihtoehtojen olla selkeitä. On tarkistettava, ettei mitään kysymystä voida tulkita monikäsitteisesti, joten yhdisteleviä sanoja kuten 'ja', 'tai' tai 'sekä' ei tulisi käyttää. Kyselyn tulisi sisältää sekä avoimia että suljettuja kysymyksiä, jotta vastaajalla on mahdollisuus kertoa mielipiteistään pidemmin tai huomauttaa lomakkeen mahdollisista puutteista. (Couper, 2008.)

Suljetuissa kysymyksissä, joissa vastausvaihtoehdot ovat valmiiksi annettuja, tulisi vaihtoehtoja olla maksimissaan viidestä seitsemään (5-7). Tämä on yleisesti sopivana pidetty määrä ihmisen käsitys- ja hahmottamiskyvylle (Vehkalahti, 2008, s.37). Vaihtoehtojen lukumäärän tulisi olla läpi lomakkeen sama, jotta vastaaja ei joudu jokaisen kysymyksen alussa pohtimaan asteikkoa uudestaan tai vastaa edellisen kysymyksen asteikon perusteella (Schwarz, Knäuper, Oyserman & Stich, 2008, s.28). Tämän perusteella valitsin kaikkiin väittämiin viisiportaisen asteikon ja lisäsin hieman erilleen ”En tiedä”- vaihtoehdon, jotta vastaajalla olisi tiedon puuttuessa mahdollisuus ohittaa myös kysymys.

Kysymysten asettelu ei saisi sisältää myöskään arvolatauksia. Esimerkiksi sosioekonomista taustaa tutkittaessa ammatit esitetään usein ”arvojärjestyksessä”. Tämä voi luoda vastaajan mielessä mielikuvan asenteellisesta tutkimuksesta ja pahimmassa tapauksessa tuottaa vastaajalle negatiivisia tunteita. Myös numerokoodit vastausvaihtoehdoille tulisi jättää pois vastaajan näkyvistä, jottei turhia arvolatauksia syntyisi. (Couper, 2008.)

Jotta vastaaja saataisiin vastaamaan myös herkkiin kysymyksiin, on hänelle tehtävä mahdollisimman selväksi tutkimuksen hyödyt ja motivoida häntä vastaamaan. Vastaaja punnitsee kyselyyn vastaamisen hyötyjä sekä haittoja ja vastatakseen hyötyjen pitää olla suuremmat. Ensimmäisenä tehtävänä on luoda kyselystä merkityksellinen myös vastaajalle selittämällä kuinka juuri hän hyötyy tutkimuksen tuloksista. Myös pienet lahjat ennen kyselyä luovat vastaajalle tunteen, että he ovat jotenkin velkaa saamistaan lahjoistaan ja vastaavat todennäköisemmin kyselyyn. Saatekirje ja itse kysely toimivat motivaattoreina. Saatekirjeessä pitäisi tulla selkeästi esille tutkimuksen tavoite, mitä tuloksilla tullaan tekemään, kuka tutkimuksen tekee, aika, joka vastaamiseen kuluu sekä osoite, johon voi ottaa yhteyttä, jos kysymyksiä ilmenee. (Lensvelt-Mulders, 2008, s.467)

Toisin kuin paperisissa kyselytutkimuksissa voidaan internet-lomakkeella asettaa joi-
tain **kysymyksiä pakollisiksi**, siten että vastaaja ei pääse etenemään lomakkeella antamatta vastausta. Tällaiset pakolliset kysymykset voivat olla haitallisia vastaajan motivaation sekä mittaustarkkuuden kannalta. Kun vastaaja pakotetaan valitsemaan jokin valmiiksi annettu vaihtoehto, on vaarana, että sopivaa vaihtoehtoa ei löydy. Tällöin vastaaja joutuu valehtelemaan ja valitsemaan itselleen väärän vaihtoehdon, tai jättämään kyselyn kesken etenemisesteen kohdatessa. Kun oikeaa vastausvaihtoehtoa ei löydy, turhautuu vastaaja usein koko kyselyyn ja jättää vastaamisen sikseen. Pakotettu valehtelu aiheuttaa siis mittausvirhettä ja valehtelulla voi olla sivuvaikutuksia myöhempien kysymysten vastausten tarkkuuteen. Hyvässä kyselyssä tulisi siis olla mahdollisimman vähän pakollisia kysymyksiä tai vastaajalle tulisi aina tarjota kysymyksen kiertävä vaihtoehto kuten ”en tiedä”. (Dillman, 2009, s.209)

Tutkijan tulisi siis valita huolella mitkä kysymykset ovat ehdottaman tärkeitä tutkimuksen kannalta ja asettaa vain nämä kysymykset pakollisiksi lomaketta laatiessa. Tässä tutkimuksessa päätin asettaa vain yhden kysymyksen pakolliseksi. Tässä kysymyksessä, neljännessä osiossa, vastaajan tulisi valita maksaako hän koulutuksestaan vai ei, minkä perusteella vastaajalle näytetään jompikumpi vaihtoehtoista sivuista (liite D osio 6a/b)

koskien rahankäyttöä, missä kysymykset on muotoiltu eri tavoin.

3.3.3 Kyselylomakkeen ja kysymysten testaus

Pahimpien virheiden huomaamiseksi ja korjaamiseksi kyselylomaketta tulisi testata ennen itse kyselyn toteuttamista. Hyviä testaajia ovat tutkimuksen kohderyhmään kuuluvat, mikä tässä tutkimuksessa tarkoittaa ulkomaalaisia tutkinto-opiskelijoita. Testauksella saadaan käsitys siitä, onko kyselyssä kysytyt kysymykset ja annetut ohjeet ymmärretty oikein, onko lomakkeessa jotain turhaa tai onko jotain olennaista jäänyt kokonaan kysymättä. (Vehkalahti, 2008, s.48.)

Lomaketta voidaan arvioida myös jälkeenpäin, jolloin on syytä kiinnittää huomiota mahdollisiin virheisiin. Näin niitä voidaan arvioida ja mahdollisia parannuksia tehdä. Tähän tarkoitukseen sopii hyvin niin puuttuvuuden kuin paradatan analysoiminenkin.

Kvalitatiivinen analyysi

AYY:n edustaja Hanna Sauli kertoi mahdollisuudesta testata lomaketta pienimuotoisesti AYY:n Tutkarikomitealla (*International Degree Students' Committee, IDSC*). Tutkarikomiteaan kuuluu vapaaehtoisia ulkomaalaisia tutkinto-opiskelijoita Aalto-yliopistosta. Kyselylomake lähetettiin Tutkarikomitean sähköpostilistalle ja Hanna Saulin kirjoittamassa lyhyessä saatekirjeessä selitettiin kommentointiavun tarve.

Kommentteja tuli yhteensä kolmelta taholta, mikä oli tarpeeksi suurimpien epäkohtien huomaamiseen. Palautteen perusteella motivaatiota mittaavaan kysymyspatteristoon lisättiin kaksi vaihtoehtoa. Kaikki palautteenantajat pitivät sosioekonomista taustaa mittaavaa osiota liian tunkeilevana ja pitkänä, joten sitä lyhennettiin ja kysymysvaihtoehtoja punnittiin uudestaan. Yhdestä palautteesta kävi myös ilmi, että aasialaisille kysymys vanhempien ammatista oli tungettelevampi kuin mitä kysymys vanhempien tulostasosta olisi ollut. Eurooppalaisesta näkökulmasta katsottuna olisin itse kokenut asian toisin. Kieliasu käytiin myös läpi saadun palautteen perusteella.

Myös Sauli itse kommentoi lomaketta useaan kertaan sen rakentamisen aikana ja oli näin suureksi avuksi. Kaikilta tahoilta saatu palaute oli hyvin arvokasta ja palautteen avulla tehdyt muutokset paransivat lomaketta huomattavasti.

Paradatan analysointi

Paradata eli tieto kyselyn prosesseista on helposti kerättävissä internet-kyselyissä automaattisten tallenteiden avulla. Paradatasta kannattaa analysoida vastaajan teknisen ympäristön tiedot kuten selain tyyppi ja versio, reaktioajat, virheet ja niiden lukumäärä, navigointikäyttäytyminen eli palaako vastaaja esimerkiksi aikaisemmille sivuille sekä viimeisin vastattu kysymys. (Dillman, 2009, s.216, 373)

E-lomakepalvelussa paradatan keräys on mahdollistettu valitsemalla tämä toiminto lomakkeen muokkausvalikosta. Palvelun pitäisi kerätä vastaajista mm. IP-osoite ja vastaamiseen kulunut aika, mutta poimimassani datassa kentät olivat täysin tyhjiä. Paradatan analyysi jää tässä tutkimuksessa siis tyngäksi, mikä on hyvin valitettavaa lomakkeen onnistumisen arvioinnin kannalta.

3.4 Aineisto

Tässä luvussa kerron aineiston keräämisestä sekä keräämisen aikana havaituista virheistä. Lisäksi kuvailen aineistoa, kerron tekemistäni muokkauksista sekä imputoinnista eli puuttuvien havaintojen korjaamisesta. Samalla pohdin kyselylomakkeen sekä kyselytutkimuksen onnistumista kerätyn aineiston perusteella.

3.4.1 Aineiston keruu

Aineisto kerättiin Aalto-yliopiston englanninkielisiltä maisteriopiskelijoilta 26.10 – 31.11.2011 verkkolomakkeen avulla. Opiskelijoiden sähköpostiin lähetettiin kutsu, josta pääsi linkkiä painamalla kyselyyn. Opiskelijoita motivoitiin osallistumaan kyselyyn arvonnalla, jossa voi voittaa 100 euron lahjakortin Stockmannille. Opiskelijoiden osoitteet saatiin AYY:n omasta rekisteristä sekä Aalto-yliopiston rekisteristä. Yliopiston hallintoimien tarkempien yhteystietojen saamiseksi laadin tarkan tutkimussuunnitelman (liite A) sekä selvityksen siitä, mihin yhteystietoja tullaan käyttämään ja kuinka tiedot tullaan hävittämään. Lisäksi vaatimuksena oli rekisteriseloste, joka tulisi olla näkyvillä myös kyselyyn vastanneille. Tietojen luovutuksesta sekä käytöstä laadittiin kirjallinen sopimus (liite B). Aalto-yliopistolta saatua listaa käytettiin vain viimeisellä muistutuskierroksella, mutta muistutuksella ei ollut enää merkittävää vaikutusta vastausprosenttiin (ks. kuvio 3.7, 30.11.2011).

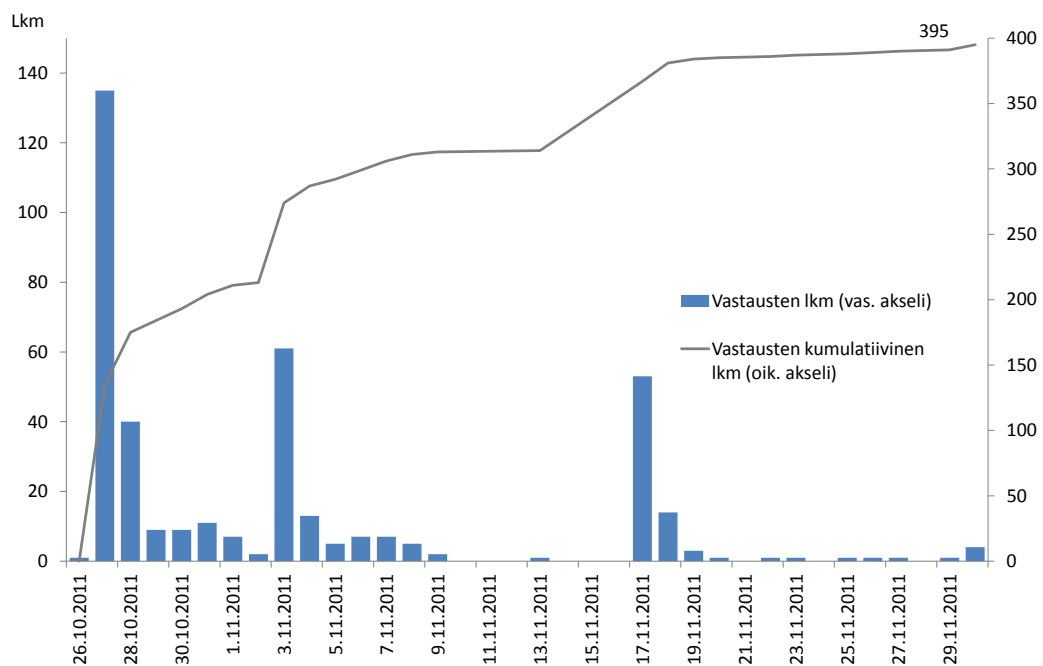
E-lomakepalvelun avulla lähetettyjen kutsujen etuna oli, että palvelu generoi jokaiselle kutsun vastaanottajalle oman käyttäjätunnuksen ja salasanan lähettäessään viestin. Salasanallinen lomake estää kohderyhmään kuulumattomien vastaukset, mahdollistaa tuplavastauksien eliminoinnin, sekä tiedonsaannin siitä, ketkä ovat jo vastanneet. Lisäksi salasanallisuus luo turvallisuuden ja luotettavuuden tunnetta vastaajissa.

Piakkoin lomakkeen lähettämisen jälkeen sain kolme ilmoitusta, etteivät vastaajat päässeet osallistumaan saatekirjeessä mainittuun arvontaan. Kävi ilmi, ettei linkki arvontaan toiminut ja näiden palauteviestien ansiosta sain korjattua virheen nopeasti.

3.4.2 Saadun aineiston kuvaus

Kysely lähetettiin 868 maisterivaiheen opiskelijalle, joiden äidinkieli oli AYY:n rekisterissä muu kuin suomi tai ruotsi, ja joilla ei ollut vaihto-oppilasstatusta. Kysymyslomakkeeseen vastasi yhteensä 395 opiskelijaa ja lopulliseksi vastausprosentiksi tuli n. 46 % (ks. kuvio 3.7). Viikon päästä ensimmäisen kutsun lähettämisestä vastausprosentti oli 25 %. Ensimmäinen muistutus lähetettiin 3.11. niille, jotka eivät olleet vielä vastanneet ja vastausprosentti nousi 36 %:iin seuraavan viikon aikana. Toisen muistutuksen (7.11.) jälkeen vastaajia oli 390 ja kolmannen 395.

Huomasin vasta aineistoon tutustuessa, että aikomani ehdollinen reititys lomakkeessa ei ollut jostain syystä toiminut ollenkaan. Näin vastaajat saivat nähtäväkseen kaksi eri versiota sivusta, jonka eri vaihtoehdot oli tarkoitettu maksaville ja ilmaiseksi opiskeleville (ks. kyselylomakkeen sivut 6a ja 6b liitteestä D). Tämä ei toisaalta haitannut kahdesta syystä. En ollut osannut ottaa huomioon Erasmus Mundus - opiskelijoiden intoa vastata kyselyyn ja tehdä ehdollistavasta kysymyksestä tarpeeksi selkeää koskemaan vain Aalto-yliopiston keräämää maksua. Toiseksi lomakkeen pitkittymisestä huolimatta vastaajat eivät olleet väsyneet, vaan lomakkeen viimeisiinkin kysymyksiin oli vastattu



Kuvio 3.7. Vastauskäyttäytyminen

suurella volyyymilla. Mainittujen sivujen ensimmäisessä versiossa (6a) oli kysymys *ExpensesA*, jonka vastausprosentti oli 73 % ja toisella vaihtoehtoiseksi tarkoitettulla sivulla (6b) muuttuja *ExpensesB*, jonka vastausprosentti oli jopa korkeampi 84 % (ks. taulukko 3.2, s. 39).

Alkuperäinen aineisto koostuu 81 muuttujasta, joista ensimmäinen on vastaajan tunnus ja toinen lomakkeen lähetyaika. Kyselylomake koostui 51 väittämästä asteikolla 1 – 5 ja 99 (99 symboloidessa 'en tiedä'-vaihtoehtoa). Kuusi kappaletta kysymyksistä oli avoimia kenttiä ja loput 26 kysymystä olivat monivalintoja. Numeeristen muuttujien tilastot ovat nähtävissä liitteestä E ja lomakkeen kaikki kysymykset liitteestä D.

Aineiston edustavuus perusjoukossa

Aineiston edustavuutta perusjoukossa voidaan arvioida vertaamalla koko perusjoukosta saatavilla olevia tietoa lomakkeen avulla kerättyihin tietoihin. AYY:n rekisteristä, josta perusjoukon yhteystiedotkin saatiin, oli mahdollisuus poimia myös tietoa opiskelijoiden kansalaisuuksista sekä iästä. Tiedot eivät kerro koko totuutta aineiston edustavuudesta, mutta antavat jonkinlaista suuntaa päätelmille.

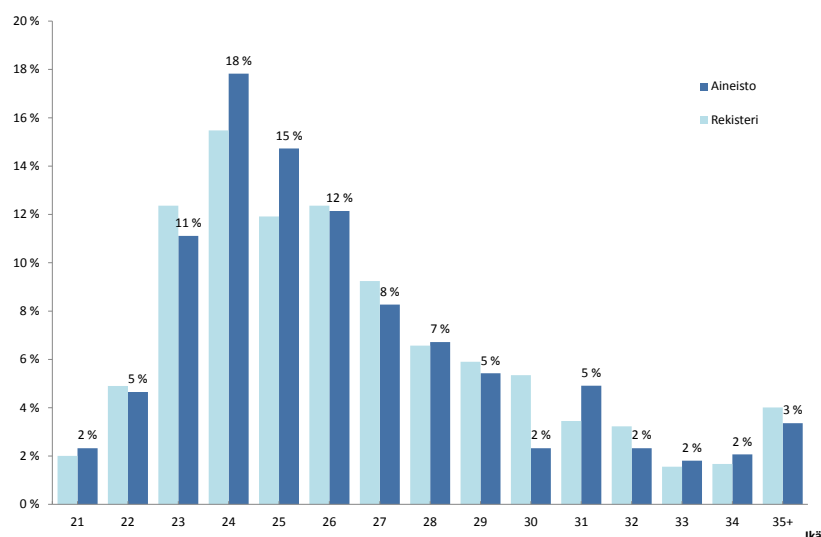
Kyselyyn tuli eniten vastauksia kiinalaisilta (13,4 %) sekä intialaisilta (11,9 %) ja nämä kansalaisuudet johtavat myös AYY:n rekisterissä määräenemmistöllään (taulukko 3.1). Näyttäisi siltä, että ainoastaan kiinalaiset ovat aliedustettuina tässä tutkimuksessa ja pakistanilaisten osuus on suurempi kuin perusjoukossa. Muiden mainittujen kansallisuuksien osalta tilanne on edustava.

Vastaajien sekä perusjoukon ikäjakaumaa kuvaa kuvio 3.8. Osuudet ovat hyvin samanlaisia, joten ikäjakaumaltaan vastauspopulaation on hyvin samanlainen. Voidaan siis

Taulukko 3.1. Maajakaumat

Maa	Aineisto	Rekisteri
Kiina	13,7 %	20,0 %
Intia	11,9 %	10,1 %
Pakistan	6,8 %	5,8 %
Venäjä	6,6 %	6,7 %
Iran	3,5 %	3,1 %
Turkki	3,0 %	2,9 %
Etelä-Korea	2,8 %	3,0 %
...

sanoa, että ikäjakaumaltaan aineisto edustaa Aalto-yliopiston englanninkielisiä maisteriopiskelijoita.

**Kuvio 3.8.** Ikäjakauma aineistossa ja perusjoukossa

Kyselyyn vastanneista 116 (29,3 %) ilmoitti aloittaneensa opiskelun vuonna 2010, 160 (40,5 %) vuonna 2011 ja loput 107 vastannutta ilmoitti aloitusvuodekseen "Other". Rekisteritietojen mukaan runsas kolmannes (36 %) on aloittanut vuonna 2010 ja jopa puolet (49 %) ovat aloittaneet vuonna 2011. Aloitusvuosien osalta aineisto ei ole siis kovin edustava.

Kyselyn vastaajista 66 % oli miehiä, 31 % naisia ja 3 % ei halunnut ilmoittaa sukupuoltaan (ks.liite E). Valitettavasti sukupuolijakaumaa ei voida vertailla AYY:n rekisteriin, mutta KOTA-tietokannan mukaan vuonna 2008 suhteet olivat 71 % ja 29 % eli naiset olisivat tämän arvion mukaan hieman yliedustettuina aineistossa.

Maksukokeiluun osallistuvista ohjelmissa opiskelevia vastasi yhteensä 86, joista 37 oli aloittanut vastausten mukaan vuonna 2011. Heistä 30 tulee EU:n ulkopuolelta, ja vastausten mukaan kuudella on A-kategorian stipendi, kymmenellä B-kategorian, kahdella C- ja yhdellä D-kategorian stipendi. Vastauksiin tulee suhtautua varauksella, sillä stipendeissä ja maksullisuudessa oli ollut paljon väärinkäsityksiä. Määrät olivat joka tapauksessa liian pieniä maksujen vaikutusten arviointiin tilastollisesti merkitsevästi.

3.4.3 Puuttuneisuus

Puuttuneisuus voi johtua kehikkovirheistä eli ali- ja ylipiteon tai huonon otannan aiheuttamasta kadosta sekä vastauskadosta. (Laaksonen, 2009)

Kehikkovirheet

Heti lomakkeen lähettämisen jälkeen sähköpostiini ilmestyi 12 virheviestiä, jotka ilmoittivat, ettei lähettämäni viesti ollut mennyt perille. Toisin sanoen ainakin 12 opiskelijan yhteystiedot olivat väärin AYY:n rekisterissä, joten syntyi joko ali- tai ylipiteoa riippuen siitä opiskelevatko kyseiset yksilöt enää. Yksi kutsun saaneista kieltäytyi henkilökohtaisesti vastaamasta kyselyyn ja yksi valitteli kyselyn sulkeutumisen jälkeen, ettei ollut maassa kyselyn aikana. Poistin muistutuskierroksilta yhteystietoja sitä mukaa, kun sain edellä mainittuja viestejä.

Vastauskato

Yksikkövastauskadon analyysi jää tyngäksi, sillä tiedossa on vain vastaamattomien lukumäärä (473) kun e-lomakepalvelun paradatan keräysominaisuus ei jostain syystä toiminut. Tietojen avulla oltaisiin voitu analysoida tarkemmin kyselylomakkeen onnistuneisuutta ja saada edes hiukan tietoa kyselyn avanneista ja keskeyttäneistä.

Taulukko 3.2. Numeeriset muuttujat, joissa puuttuvuus oli korkea

Muuttuja	Label	Puuttuvia lkm	Puuttuvuus %
V_6	The differences in quality between the free and chargeable master degree programs in Aalto are high	222	56 %
Aalto_14	Other	192	49 %
V_3	Paying and non-paying students are treated equally in Aalto	150	38 %
FI_14	Other	133	34 %
ExpensesA	Estimation of your monthly expenses in Finland in euros (incl. Rent, food, transportation etc.)	108	27 %
ExpensesB	Estimation of your monthly expenses in Finland in euros (incl. Rent, food, transportation etc. BUT excluding tuition fees)	63	16 %
Aalto_11	Scholarship	47	12 %

Osittaista puuttuvuutta kysymyksittäin tai vastaajittain eli erävastauskatoa syntyy siitä, että kohderyhmän edustaja vastaa kyselyyn puutteellisesti (Laaksonen, 2009). Muuttujittain puuttuneisuutta esiintyi luonnollisesti etenkin kysymyksissä A_14 ja FI_14, jotka olivat avoimia kenttiä ja joissa pyydettiin ilmoittamaan muita syitä. Kysymys V_6 oli jälkeinpäin ajateltuna aivan liian pitkä sekä monimutkainen. Kysymykseen oli vastattu huonosti varmasti tämän vuoksi (taulukko 3.2). Myös kysymys V_3 oli vastaajille vaikea ja vastaamatta jätti 38 %. Muita runsaan puuttuvuuden kysymyksiä olivat rahankäyttöä koskevat kysymykset sekä stipendin vaikutus opiskelupaikan houkuttelevuuteen, sillä suurin osa vastaajista ei varmasti tiennyt edes stipendien olemassaolosta Aalto-yliopistossa.

Taulukko 3.3. Vastausaktiivisuus

Erävastauskato/vastaaaja	Opiskelijoita
Alle 10 %	37
10 – 19 %	255
20 – 29 %	83
30 – 39 %	14
Yli 40 %	6

Kaikki kyselyyn vastanneet olivat täyttäneet lomakkeen ahkerasti ja vain neljällä vastaajalla oli yli 40 tyhjäksi jäänyttä kysymystä. Suurin osa oli jättänyt vastaamatta n. 10 – 19 % (8 – 15 kpl) kysymyksistä. Vastaamattomuus näyttäisi korreloivan ainoastaan negatiivisesti syntymävuoden kanssa (-0,34), joten nuoremmat ovat vastanneet vanhempia ahkerammin.

3.4.4 Aineiston muokkaus

Aineiston muokkaus käsittää alkuperäisten muuttujien muunnokset, virheiden korjaukset sekä puuttuvien havaintojen korvaamisen.

Aineiston tarkistus ja muunnokset

Aineiston tarkistuksen tavoitteena on varmistaa aineiston laatu eli että aineisto on siinä muodossa kuin sen on tarkoitus olla (Groves ym., 2004). Tarkistuksessa kannattaa käydä läpi että vastaukset ovat halutulla asteikolla ja halutulla välillä ja että vastaukset ovat mahdollisia sekä loogisia.

Aineistoon tutustuttuani huomasin muutaman poikkeavan havainnon, jotka eivät loogisesti voineet olla totta. Korjasin kyseiset vastaukset oikeammiksi sopivimmalla nähdyllä tavalla. Esimerkiksi yhden vastaajan syntymäaika oli 1800-luvulla, joten korjasin sen 1900-luvulle. Lisäksi monet Erasmus Mundus – ohjelmiin kuuluvat olivat ymmärtäneet stipendikysymyksen väärin, eivätkä he löytäneet heille tarkoitettua kohtaa koulutusohjelmalistasta, vaan kirjoittivat omansa kohtaan 'Other'.

Muutin 'En tiedä' - vaihtoehdot puuttuviksi analyysijä varten, sillä ne eivät kuulu lähtökohtaisesti numeeriselle asteikolle, sillä ne mittaavat eri asiaa kuin kysymystä (Vehkalahti, 2008, s.26, 83). Lisäksi huomasin, että joitain alkuperäisiä muuttujia on syytä muuntaa tai koodata uudelleen ja laskin vastaajien iän syntymävuosista ja luokittelin vielä iän viiteen luokkaan. Lisäsin aineistoon kansallisuuksien perusteella maanosat, kuuluvuuden EU- tai ETA- alueeseen ja liitin aineistoon Maailmanpankin luokittelun opiskelijoiden kotimaan tuloluokasta⁵. Lisäsin myös ohjelmien perusteella Aalto-yliopiston koulut ja jaoin järkevyystarkastelun perusteella vastaajat lukukausimaksukokeilussa oleviin.

Aineiston imputointi

Imputoinnilla tarkoitetaan puuttuvien tai epäkelvojen havaintojen täyttämistä tai korvaamista jollakin korvikearvolla. Imputoinnilla pyritään parantamaan estimoinnin laatua ja

⁵<http://data.worldbank.org/about/country-classifications>

näin parantamaan tilastollista analyysiä. Jotta laatu paranisi imputoimalla, on imputointitoiminnon oltava riittävän hyvä, sillä huonolla imputoinnilla aineiston laatu voi jopa heiketä. (Laaksonen, 2009, s. 174)

Imputointia tehdessä on muistettava, ettei mikään paikkaustapa aidosti korvaa puuttuvaksi jäänyttä tietoa ja tuloksiin tulisi suhtautua aina varauksella (Vehkalahti, 2008, s.81). Mitä enemmän puuttuvuutta esiintyy sitä enemmän tulosten tulkinnassa on epävarmuuksia (Vehkalahti, 2008).

Imputointimenetelmät olettavat puuttuvuuden olevan satunnaista, mikä ei aina ole totta (Vehkalahti, 2008, s.81). Käsiteltävässä aineistossa systemaattista puuttuvuutta esiintyi ainakin edellä mainituissa huonoissa tai täydentävissä kysymyksissä (ks. taulukko 3.2). Nämä kysymykset päätin jättää pois kokonaan analyyseistä. Oletan muiden kysymysten kohdalla puuttuneisuuden nyt satunnaiseksi.

Imputointi oli tarpeen tämän aineiston kohdalla, sillä analysointimenetelmien tulokset jäivät alkuperäisellä aineistolla laihoiksi mm. siksi, että faktorianalyysiin tarvittava korrelaatiomatriisi supistui liian pieneksi.

Imputointimenetelmäksi valittiin regressioimputointi, jossa puuttuneita arvoja estimoidaan usean muuttujan lineaarisella regressioanalyysillä. Tämä menetelmä on hyvin yksinkertainen ja se valittiin siksi, että tutkimus toimii pilottitutkimuksena, eikä tuloksia pyritä edes käyttämään estimoimaan perusjoukkoa. Suoritin imputoinnin SPSS:n puuttuneisuusanalyysi-proseduurin avulla. Proseduurin syntax löytyy liitteistä (liite G).

4 Monimuuttujamenetelmät

Monimuuttujamenetelmissä käsitellään useita muuttujia samanaikaisesti ja menetelmien tarkoituksena on havaintomatriisin eli aineiston informaation tiivistäminen jonkin mallin avulla. Monimuuttujamenetelmillä saadaan siis tiivistettyä aineistoa ja näin selkeytettyä monilla muuttujilla kuvattua aihetta. Muuttujien ja alkioiden vaihtelusta pyritään karSIMaan pois satunnainen osuus ja paljastamaan taustalla piilevät rakenteet. (Mustonen, 1995, s.iii.)

Tässä työssä esittelen monimuuttujamenetelmistä faktorianalyysiä, jolla saadaan sopivasti tiivistettyä aineistoa, sekä ryhmittelyanalyysiä, jolla aineistosta pyritään löytämään homogeenisia ryhmiä.

4.1 Faktorianalyysi

Faktorianalyysissä on kyse mahdollisimman vähäulotteisen piilorakenteen löytämisestä muuttujien korrelaatioiden avulla (Mustonen, 1995, s. 75). *Faktorit* koostuvat alkuperäisistä muuttujista, jotka yhdessä selittävät muuten vaikeasti tutkittavaa isompaa kokonaisuutta (Mustonen, 1995, s. 75). Faktorianalyysi tiivistää siis tietoa alkuperäisestä aineistosta.

Faktorianalyysissä on kaksi toisistaan poikkeavaa lähestymistapaa. Ensimmäinen eli *eksploratiivinen* faktorianalyysi on aineistolähtöinen; siinä pyritään etsimään muuttujien joukosta faktoreita, jotka pystyvät selittämään havaittujen muuttujien vaihtelua ilman tutkijan ennakkotietoja mahdollisesta ilmiöstä. Löydetyt faktorit toimivat tulkinnan ja päättelyn perustana. *Konfirmatorisessa* faktorianalyysissä tutkijalla on jo etukäteen hyvät tiedot ilmiön rakenteesta ja taustoista. Hänellä on myös mielessään valmis kuva mahdollisesta faktorirakenteesta, ja analyysin tehtävänä on joko varmistaa tai kumota tämä käsitys. (KvantiMOTV, 2004.)

4.1.1 Faktorianalyysimalli

Mallin esittelyssä olen käyttänyt samoja merkintöjä kuin Mustonen 1995, (s.75–94).

Yleisesti havaitut muuttujat X_1, X_2, \dots, X_p oletetaan faktorianalyysissä normaalisti jakautuneiksi, eli $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ja $\boldsymbol{\Sigma} > \mathbf{0}$. Muuttujilla tarkoitetaan kyselytutkimuksessa kysymyksiä, joiden arvot muuttuvat vastaajien antamien vastausten mukaan. Vehkalahden (2008, s.95) mukaan oletus ei kuitenkaan usein päde käytännössä, sillä harvan mielpidekysymyksen vastaukset ovat jakautuneet normaalisti. Faktorianalyysi sietää kuitenkin poikkeamia hyvin. (Mustonen, 1995)

Faktorianalyysi perustuu korrelaatiokertoimiin joiden laskeminen edellyttää, että muuttujien \mathbf{X} välillä on selkeitä lineaarisia riippuvuuksia. Tästä seuraa, että ainakin luo-

kitteluasteiset muuttujat jäävät tämän menetelmän tarkastelun ulkopuolelle. Muuttujat X_1, X_2, \dots, X_p esitetään seuraavan mallin avulla:

$$X_i = \mu_i + a_{i1}F_1 + a_{i2}F_2 + \dots + a_{ik}F_k + U_i, i = 1, 2, \dots, p$$

eli

$$(4.1) \quad \mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{F} + \mathbf{U},$$

jossa $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_p)$ ovat kyseisten muuttujien X_i odotusarvoja, joilla ei ole faktorianalyysin kannalta yleensä merkitystä, mutta ne on hyvä merkitä näkyviin.

Matriisi A on $p \times k$ faktori- eli latausmatriisi. Faktorilatauksella tarkoitetaan kunkin muuttujan ja sille määritellyn faktorin välistä korrelaatiokerrointa, joka vaihtelee $[-1, 1]$. Luku kertoo, kuinka paljon muuttujan vaihtelusta voidaan selittää faktorin avulla. Faktorilatauksista a_{ij} muodostuu faktorimatriisi **A**, joka on täysasteinen $r(\mathbf{A}) = k$ eli matriisin **A** aste on faktorien lukumäärä.

Muuttujien vaihtelun ajatellaan jakautuvan kahteen osaan: yhteisvaihteluun, josta olemme kiinnostuneita, ja ominaisvaihteluun, joka voidaan luokitella virhetermiksi tai kohinaksi. Yhteisvaihtelua kuvataan faktoreilla $\mathbf{F} = (F_1, F_2, \dots, F_k)$, jotka on painotettu mallissa faktorilatauksilla. Yhteisfaktoreita on vähemmän kuin alkuperäisiä muuttujia, eli $k < p$, sillä faktorianalyysin ideana on aineiston tiivistäminen. $\mathbf{U} = (U_1, U_2, \dots, U_p)$ ovat taas ominaisfaktoreita, jotka tulkittiin siis mittausvirheeksi. Oletetaan, että myös vaihtelutermit ovat jakautuneet normaalisti eli $\mathbf{F} \sim N(\mathbf{0}, \boldsymbol{\Phi})$ ja $\mathbf{U} \sim N(\mathbf{0}, \boldsymbol{\Psi}^2)$, jossa $\boldsymbol{\Psi}^2$ on alkioiden $\psi_1^2, \psi_2^2, \dots, \psi_p^2$ lävistämämatriisi. Lisäksi oletetaan, että ominaisfaktorit **U** ovat riippumattomia yhteisfaktoreista **F**, jolloin niiden välinen korrelaatiokerroin on nolla.

Kovarianssimatriisi saadaan laskemalla $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = E[(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))'] = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = \mathbf{A}E(\mathbf{F}\mathbf{F}')\mathbf{A}' + E(\mathbf{U}\mathbf{U}')$, jolloin tuloksena on yhtälö

$$\boldsymbol{\Sigma} = \mathbf{A}\boldsymbol{\Phi}\mathbf{A}' + \boldsymbol{\Psi}^2,$$

jota sanotaan faktorianalyysin perusyhtälöksi. Kun oletetaan vielä, että faktorit eivät korreloi keskenään ($\boldsymbol{\Phi} = \mathbf{I}$), päädytään yksinkertaisempaan muotoon

$$\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}' + \boldsymbol{\Psi}^2.$$

4.1.2 Suurimman uskottavuuden faktorointi

Estimointitapoja on monia, mutta yksi keskeisimmistä on suurimman uskottavuuden menetelmä, sillä se on riippumaton muuttujien mittayksiköistä. Oletetaan, että havaitut X_1, X_2, \dots, X_n ovat riippumattomia ja satunnaisotos multinormaalijakaumasta $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, jossa kovarianssimatriisi on muotoa $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}' + \boldsymbol{\Psi}^2$, kun faktorit oletetaan keskenään korreloimattomiksi eli $\boldsymbol{\Phi} = \mathbf{I}$. Matriisi **A** on $p \times r$ -faktorimatriisi ja $\boldsymbol{\Psi}^2 = \text{diag}(\psi_1^2, \psi_2^2, \dots, \psi_p^2)$. Muuttujien **X** yhteistiheysfunktio on muotoa

$$L(x; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{j=1}^n (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(x_j - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(x_j - \boldsymbol{\mu})\right]$$

$$= (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left[-\frac{1}{2} \sum_{j=1}^n (x_j - \mu)' \Sigma^{-1} (x_j - \mu)\right].$$

Funktiossa μ :ta eli havaittujen muuttujien odotusarvoa pidetään lähinnä kiusaparametrina. Tästä kiusaparametrasta halutaankin päästä eroon ennen funktion maksimointia, joten se voidaan korvata esimerkiksi parametrin omalla suurimman uskottavuuden estimaatilla \bar{x} .

Logaritmoitu uskottavuusfunktio on muotoa

$$l(\hat{\mu}, \Sigma) = -\frac{1}{2} [np \log(2\pi) - np \log |\Sigma| - \sum_{j=1}^n (x_j - \bar{x})' \Sigma^{-1} (x_j - \bar{x})].$$

Faktorianalyysin tuloksen saavuttamiseksi funktio on maksimoitava parametrien \mathbf{A} ja Ψ suhteen, kun $\Sigma = \mathbf{A}\mathbf{A}' + \Psi$. Ratkaisu ei ole yksikäsitteinen ja se saadaan vain numeerisesti laskemalla. Jöreskogin (2004) mukaan parempi tapa on minimoida log-uskottavuusfunktio muistuttava funktio

$$F_{ML}(\mathbf{A}, \Psi) = \log |\Sigma| + \text{tr}(\mathbf{S}\Sigma^{-1}) - \log |\mathbf{S}| - p,$$

jossa \mathbf{S} on otoksesta laskettu kovarianssimatriisi ja edelleen $\Sigma = \mathbf{A}\mathbf{A}' + \Psi^2$.

4.1.3 Rotaatiomenetelmät

Faktorit \mathbf{F} eivät kuitenkaan määräydy yksikäsitteisesti perusyhtälöstä ja oletuksista, vaan faktoreiden määrittämiseen käytetään rotaatiomahdollisuutta. Alkuperäistä ratkaisua voi siis kiertää faktorien virittämässä koordinaatistossa (faktoriakselien kierto) niin, että tulos näkyy tulokinnan kannalta mahdollisimman yksinkertaisessa muodossa. Kierto tapahtuu valitsemalla matriisi \mathbf{T} niin, että $\mathbf{F}^* = \mathbf{T}'\mathbf{F}$, jossa \mathbf{T} on säännöllinen $r \times r$ -matriisi, jolloin $\mathbf{F}^* \sim N(\mathbf{0}, \mathbf{T}'\Phi\mathbf{T})$ ja $\mathbf{A}^* = \mathbf{A}(\mathbf{T}^{-1})'$. Näin saadaan muuttujille \mathbf{X} esitys

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}^* + \mathbf{F}^* + \mathbf{U}$$

ja perusyhtälö tulee muotoon

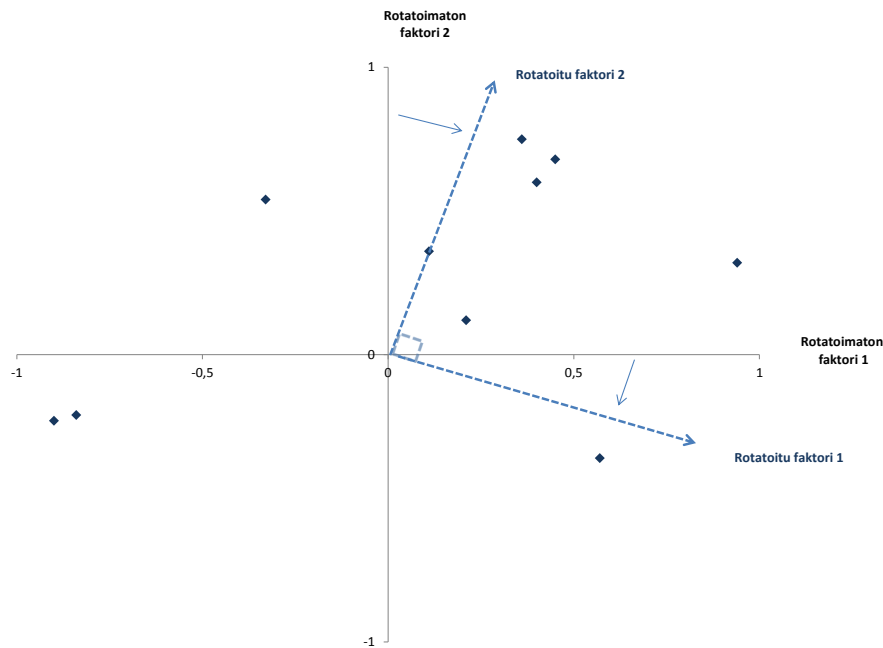
$$\Sigma = \mathbf{A}\Phi\mathbf{A}' + \Psi^2 = \mathbf{A}^*\Phi^*\mathbf{A}^{*'} + \Psi^2.$$

Jokainen \mathbf{F} :n rotaatio ja sitä vastaava faktorimatriisi on myös mahdollinen ratkaisu. Rotaation ongelmana on siis mahdollisuus valita mikä tahansa ratkaisu. Tutkijan ennakkotiedot ilmiöstä ovat tässä tarpeen, sillä analyysin tekijän pitää päättää r eli faktoreiden määrä sekä laskea matriisit Ψ^2 , \mathbf{A} ja \mathbf{T} tietokoneen avulla. Niiden määrittämisestä kutsutaan faktoroinniksi ja rotatoinniksi. (Mustonen, 1995, s.85)

Rotaatiomenetelmät jaetaan kahteen pääluokkaan: suorakulmarotaatiomenetelmiin (*orthogonal rotation*) ja vinokulmarotaatiomenetelmiin (*oblique rotation*). Ensimmäinen menetelmä tuottaa sellaisia faktoreita, jotka eivät korreloi keskenään eli ovat riippumattomia, toinen tuottaa taas sellaisia faktoreita, jotka voivat korreloida keskenään. Korreloimattomat faktorit ovat tulokinnan kannalta helpompia, sillä ne näyttäytyvät kuvassa suorassa kulmassa toisiinsa nähden kuten tavallisen koordinaatiston x- ja y-akselit. (Nummenmaa, 2004, s.346)

Graafinen rotaatio

Graafista rotaatiota käytettiin ennen tietokoneiden laskentakapasiteetin äkillistä nousua, ja nykyäänkin graafinen esitys on hyvä keino havainnollistaa muuttujien latauksia eri faktoreille (Mustonen, 1996). Menetelmässä piirretään faktoriavaruuden kaksiulotteisia projektioita eri faktoriparien suhteen, minkä jälkeen faktorimatriisi lasketaan uudelleen valitun rotaation mukaisesti. Valitaan esimerkiksi faktorit F1 ja F2, piirretään ne niin, että F1 on kuin perinteinen x-akseli ja F2 kuin y-akseli (esim. kuva 4.1). Tämän jälkeen lasketaan malliin valituille muuttujille faktorilataukset ja piirretään ne koordinaatistoon origosta lähteviksi vektoreiksi. Eri faktoripareja valitaan niin kauan, kunnes löydetään paras eli selkein ratkaisu.



Kuvio 4.1. Faktoriavaruuden ortogonaalinen rotatointi

Analyttiset rotaatiot

Analyttiset rotaatiot perustuvat rotatoidusta faktorimatriisista $\mathbf{A}^* = \mathbf{A}(\mathbf{T}^{-1})'$ riippuvaan mittaan $m(\mathbf{A}^*)$, joka maksimoidaan tai minimoidaan matriisin \mathbf{T} suhteen.

Esimerkkinä analyttisestä ortogonaalisesta rotaatiosta mainittakoon suosituin, *Varimax*-menetelmä. Menetelmässä maksimoidaan faktorilatausten neliöitten sarakkeittain (eli faktoreittain) laskettujen varianssien summa.

$$m(\mathbf{A}^*) = p^2 \sum_{j=1}^r v_j^2,$$

missä

$$v_j^2 = \frac{1}{p} \sum_{i=1}^p [a_{ij}^* - \frac{1}{p} \sum_{k=1}^p a_{kj}^*]^2.$$

Menetelmä tuottaa latausmatriisin, jossa joidenkin muuttujien lataukset faktoreiden suhteen ovat korkeat ja toisten matalat. Käytännössä ratkaisu joudutaan estimään iteratiivisesti toistamalla kaksiulotteisia kiertoja kuten graafisessa rotaatiossa.

4.1.4 Faktoripisteet

Faktoripisteillä tarkoitetaan faktoreita kuvaavia havaintokohtaisia lukuja. Yhdessä faktoripisteet kuvaavat faktoreiden edustamien ulottuvuuksien jakautumista aineistossa. Faktoripisteet ovat käyttökelpoisia jatkoanalyysissä, sillä niiden perusteella havaintoja voidaan esimerkiksi ryhmitellä tai tarkastella muuten tiivistettyä tietoa. Faktoripisteiden käyttö edellyttää, että faktorit pystytään nimeämään eli ne ovat hyviä sekä selkeitä uusia muuttujia. (Vehkalahti, 2008, s.109.)

Faktoripisteitä voidaan laskea ainakin kahdella tavalla: painotetulla PNS-menetelmällä sekä regressiomenetelmällä.

Ensimmäisessä, eli **PNS-menetelmässä**, oletetaan, että vektori \mathbf{x} on faktorimallin 4.1 mukainen multinormaalinen satunnaisvektori. Oletetaan, että matriisit \mathbf{A} ja $\mathbf{\Psi}$ ovat tunnettuja ja että $\boldsymbol{\mu} = \mathbf{0}$ ja olkoon \mathbf{f} ($p \times k$) yhteisvaihtelua kuvaava faktorivektori, joka on jakautunut $\mathbf{x} \sim N_p(\mathbf{A}\mathbf{f}, \mathbf{\Psi})$. (Mardia, Kent & Bibby, 1989, s.273.)

Log-uskottavuusfunktioksi saadaan

$$l(\mathbf{x}; \mathbf{f}) = -\frac{1}{2}(\mathbf{x} - \mathbf{A}\mathbf{f})'\mathbf{\Psi}^{-1}(\mathbf{x} - \mathbf{A}\mathbf{f}) - \frac{1}{2}\log |2\pi\mathbf{\Psi}|.$$

Log-uskottavuusfunktion maksimi löytyy sen derivaatan nollakohdista, joten

$$\begin{aligned}\frac{\partial l}{\partial \mathbf{f}} &= -\frac{1}{2}(-2\mathbf{A}'\mathbf{\Psi}^{-1}\mathbf{x} + 2\mathbf{A}'\mathbf{\Psi}^{-1}\mathbf{A}\mathbf{f}). \\ &= \mathbf{A}'\mathbf{\Psi}^{-1}(\mathbf{x} - \mathbf{A}\mathbf{f}).\end{aligned}$$

ja asettamalla yhtälö nolaksi saadaan

$$\mathbf{A}'\mathbf{\Psi}^{-1}\mathbf{x} = \mathbf{A}'\mathbf{\Psi}^{-1}\mathbf{A}\mathbf{f}.$$

$$(4.2) \quad \hat{\mathbf{f}} = (\mathbf{A}'\mathbf{\Psi}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{\Psi}^{-1}\mathbf{x},$$

joka tunnetaan myös Bartlettin faktoripistemääränä. (Mardia ym., 1989, s.274.)

Käytännössä matriisit \mathbf{A} , $\mathbf{\Psi}$ ja $\boldsymbol{\mu}$ eivät ole tunnettuja vaan niiden estimaatit lasketaan aineistosta. Korvaamalla lausekkeessa 4.2 matriisit estimaateilla, saadaan siis painotetut PNS-estimaatit (Mellin, 2009, s.16).

Regressiomenetelmässä pyritään löytämään kerroinmatriisi \mathbf{K} , joka toteuttaisi niin hyvin kuin mahdollista yhtälön $\mathbf{f} = \mathbf{K}(\mathbf{x} - \boldsymbol{\mu})$, jossa \mathbf{K} on $r \times p$ -matriisi (Mellin, 2009). Muodostamalla neliösumma

$$\varphi(\mathbf{K}) = E[(\mathbf{f} - \mathbf{K}(\mathbf{x} - \boldsymbol{\mu}))'(\mathbf{f} - \mathbf{K}(\mathbf{x} - \boldsymbol{\mu}))]$$

ja laskemalla kerroinmatriisi

$$(4.3) \quad \mathbf{K} = \mathbf{A}'\boldsymbol{\Sigma}^{-1},$$

joka minimoi neliösumman $\varphi(\mathbf{K})$, saadaan yhtälö $\mathbf{f} = \mathbf{A}'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$. Lause voidaan kirjoittaa myös muotoon

$$\mathbf{f} = (\mathbf{A}'\mathbf{\Psi}^{-1}\mathbf{A} + \mathbf{I})^{-1}\mathbf{A}'\mathbf{\Psi}^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

Korvaamalla matriisit estimaateillaan, kuten edellä, saadaan faktoripisteiden regressio-estimaatit. (Mellin, 2009.)

4.2 Ryhmittelyanalyysi

Ryhmittelyanalyysillä pyritään löytämään aineistosta havaintojen tai muuttujien ryhmiä (Johnson & Wichern, 2007, s.671). Menetelmässä mitataan havaintojen etäisyyksiä valittujen muuttujien suhteen, jotta löydetäisiin ryhmien väliset yhteydet sekä eroavaisuudet. Ryhmien muodostamiseksi käytetään laskennallista algoritmiä (esim. kuvio 4.2). Ryhmittelyn ideana on, että samaan ryhmään kuuluvien havaintojen pitäisi olla mahdollisimman samanlaisia ja toisaalta eri ryhmiin kuuluvien havaintojen pitäisi olla hyvin erilaisia. Tämä tarkoittaa, että ryhmien pitäisi olla sisäisesti homogeenisia ja ulkoisesti erilaisia, joten ryhmien välisten erojen pitäisi olla mahdollisimman selväpiirteisiä. (Mellin, 2007, s.5.)

Ryhmittelyanalyysin ideana on siis löytää aineistosta alaryhmiä, joiden jäsenet ovat keskenään samanlaisia ja muihin ryhmiin nähden erilaisia. Analyysimenetelmän teoreettinen pohja ei ole kovin luja vaan menetelmä on hyvin eksploraatiivinen eli kokeileva. Ryhmistä ei ole siis välttämättä etukäteistietoa ja parhaaseen tulokseen päästää kokeilemalla eri vaihtoehtoja. Paras ratkaisu on löydetty kun analyysiä toistettaessa saadaan suurin piirtein samat ryhmät, tai kun aineiston puolitus tuottaa samat tulokset. (Nummenmaa, 2004, s.363.)

Koko menetelmään liittyy siis paljon subjektiivisuutta, ja menetelmäkombinaation valinta riippuu tutkijan tietämyksestä analysoitavasta aiheesta, muuttujien mitta-asteikosta ja etäisyysmitan valinnasta. Lisäksi eri ryhmittelymenetelmät ovat erittäin herkkiä poikkeavien havaintojen (*outliers*) suhteen. (Johnson & Wichern, 2007, s.695.)

Parhaimman tuloksen saamiseksi tulisi ryhmittelyyn käytettyjen muuttujien vaihdella riittävästi aineistossa eikä sisältää poikkeavia havaintoja. Kaikkien muuttujien tulisi olla skaalattu samalle asteikolle (standardointi), jotta välttyään isompia arvoja saavien muuttujien painottumiselta. (Nummenmaa, 2004.)

Kuten edellä todettiin, ryhmittelyanalyysi on heuristinen menetelmä. Näin tulosten reabiliteettiin ja validiteettiin pitäisi kiinnittää huomiota ennen yleistävien päätelmien tekoa. Riittäviä tarkasteluja tulee tehdä toistamalla analyysi erilaisilla etäisyysmitoilla sekä erilaisilla ryhmittelymenetelmillä. Aineiston kahtiajako tai satunnaisotoksen ottaminen ja tulosten vertailu voi olla myös tarpeen. Ei-hierarkisissa menetelmissä aineiston järjestystä kannattaa muuttaa eri muuttujien suhteen ja vertailla näin saatuja tuloksia. (Malhotra & Birks, 2007, s.684.)

Ryhmien lukumäärän valinnalle ei ole mitään eksaktia päätäntätapaa vaan luvun valinta riippuu tilanteesta. Jos ryhmien lukumäärästä on teoreettista aavistusta tai tietoa, tulisi tätä käyttää hyväksi. Teoreettisen taustan puuttuessa voidaan keskittyä tilasto-ohjelmien tulosteiden tulkintaan. Hierarkisissa menetelmissä tulisi keskittyä etäisyyksiin, joilla ryhmät on muodostettu. Ei-hierarkisissa menetelmissä ryhmien sisäisen ja välisen varianssin suhde on tärkeä mittari ja luvut voidaan piirtää kuvaajaksi helpottamaan tulkintaa. Myös ryhmien suhteelliseen kokoon tulisi kiinnittää huomiota, esimerkiksi yhden tai kahden alkion ryhmät eivät usein ole järkeviä. (Malhotra & Birks, 2007, s.608.)

Ei ole siis olemassa yhtä oikeaa ryhmittelymenetelmää, vaan tilanteesta riippuen on löydettävä sopivin vaihtoehto tulosten analysoimiseen ja useimmiten tutkimuksen teko vaatii useamman klusterointitavan kokeilemista (Malhotra & Birks, 2007). Esittelen seuraavaksi menetelmistä kolme, joita tulen käyttämään käytännön analyysissä luvussa 5.

Analyysimenetelmän vaiheet ovat siis etäisyysmitan valitseminen, josta kerron seu-

raavaksi, sekä ryhmittelytekniikan valinta, mitkä jaetaan hierarkisiin ja ei-hierarkisiin menetelmiin, joista etäisyysmittojen jälkeen.

4.2.1 Etäisyys- ja erilaisuusmitat

Ryhmittelyanalyysin ideana on havaintojen samankaltaisuuden tai erilaisuuden mittaaminen niiden etäisyyksien avulla (Mellin, 2007). Etäisyysmittoja on monia perustuen sekä jatkuviin että dikotomisiin muuttujiin (Mardia ym., 1989, s.376–378). Jatkuville muuttujille käytetään etäisyysmittoja ja dikotomisille muuttujille erilaisuusmittoja. Esitelen niistä muutaman. Käytän kappaleessa samoja notaatioita kuin Mellin (2007), Mardia, Kent & Bibby (1989) ja Johnson & Wichern (2007).

Etäisyysmitta d täyttää seuraavat ehdot:

$$d(i, j) \geq 0 \text{ kaikilla } i, j,$$

$$d(i, i) = 0,$$

$$d(i, j) = d(j, i) \text{ kaikilla } i, j.$$

(Mardia ym., 1989, s.376.)

Etäisyysmitat jatkuville muuttujille

Olkoon \mathbf{X} $n \times p$ -matriisi, jonka rivit ovat $\mathbf{x}'_1, \dots, \mathbf{x}'_n$. Olkoon pisteet, joiden välimatkaa mitataan, muotoa $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ ja $\mathbf{x}_j = (x_{j1}, \dots, x_{jp})$.

Matriisin \mathbf{X} pisteiden \mathbf{x}_i ja \mathbf{x}_j välinen **Minkowski-etäisyys** on muotoa

$$D_{ij} = \left(\sum_{k=1}^p |x_{ik} - x_{jk}|^m \right)^{1/m}.$$

Kun $m = 1$ puhutaan Manhattan- tai City Block -etäisyydestä ja kun $m = 2$ on etäisyys tavanomainen euklidinen etäisyys. **Euklidinen etäisyys** kahden pisteen \mathbf{x}_i ja \mathbf{x}_j :n välillä on määritelty seuraavasti.

$$D_{ij} = \left(\sum_{k=1}^p |x_{ik} - x_{jk}|^2 \right)^{1/2},$$

joka voidaan myös kirjoittaa muodossa

$$= \sqrt{(\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j)}.$$

Tilastollinen etäisyys (*Statistical distance*) SD_{ij} saadaan seuraavasta määritelmästä

$$SD_{ij}^2 = \sum_{k=1}^p \left(\frac{x_{ik} - x_{jk}}{s_k} \right)^2,$$

jossa s_k on muuttujan x_k , $k = 1, \dots, p$ havaituista arvoista x_{tk} , $t = 1, \dots, n$ laskettu otoshajonta

$$s_k^2 = \frac{1}{n-1} \sum_{t=1}^n (x_{tk} - \bar{x}_k).$$

Yhtälössä \bar{x}_k on muuttujan x_k havaittujen arvojen otoskeskiarvo.

Mahalanobis-etäisyysmitassa käytetään hyväksi havaituista arvoista laskettua kovarianssimatriisia. Etäisyysmitta on joissain tilanteissa parempi kui tavallinen Euklidinen etäisyys, sillä se ottaa huomioon muuttujien keskinäisen riippuvuuden eikä se ole riippuvainen mitta-asteikosta (Mustonen, 1995, s.41) Etäisyys MD_{ij} määritellään seuraavasti:

$$MD_{ij} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)' \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_j)},$$

jossa \mathbf{S} on havaituista arvoista laskettu otoskovarianssimatriisi. Matriisi \mathbf{S} on muotoa

$$[\mathbf{S}]_{ij} = \frac{1}{n-1} \sum_{t=1}^n (x_{ti} - \bar{x}_i)(x_{tj} - \bar{x}_j)$$

, kun $i, j = 1, \dots, p$.

Erilaisuusmitat dikotomisille muuttujille

Kun muuttujan arvoa ei voida esittää järkevästi p -ulotteisena mittana siirrytään käyttämään dikotomista asteikkoa, joka kertoo jonkun ominaisuuden läsnäolosta (1) tai poisolosta (0) (Johnson & Wichern, 2007, s. 674).

Olkoon x_{ij} j :nen binäärinen muuttujan i :nen havainnon tulos (0 tai 1) ja vastaavasti x_{kj} j :nen binäärinen muuttujan k :nen havainnon tulos kun $j = 1, 2, \dots, p$. Etäisyyttä voitaisiin mitata havaintojen x_{ij} ja x_{kj} euklidisella etäisyydellä ja laskea näin yhteenso-pimattomien parien lukumäärä. Tämän menetelmän ongelma on se että 1-1 pareja sekä 0-0 pareja käsitellään samanarvoisina, vaikka usein ominaisuuden esiintyminen (esim. osaa italiaa) on vahvempi merkki yhteydestä kuin ominaisuuden puuttuminen (ei osaa italiaa). Vaihtoehtojen painottamisen sallimiseksi käytetään ristiintaulukoitiin perustu-via mittoja. (Johnson & Wichern, 2007, s.675.)

Taulukko 4.1. Binääristen muuttujien ristiintaulukko

		k:s havainto		Tulos
		1	0	
i:s havainto	1	a	b	a+b
	0	c	d	c+d
Tulos		a+c	b+d	p= a+b+c+d

Lähde: (Johnson & Wichern, 2007, s.675)

Binääristen muuttujien ristiintaulukko voidaan esittää taulukon 4.1 lailla. Taulukossa a edustaa 1-1 parien lukumäärää, b 1-0 parien, c 0-1 parien ja d 0-0 parien lukumäärää. Näiden lukumäärien tai frekvenssien kautta voidaan määrittää erilaisuusmittoja, joista esittelen kolme.

Kun painotus 1-1 ja 0-0 havainnoille on sama voidaan samanlaisuusmitta esittää muodossa

$$\frac{a+d}{p} = \frac{a+d}{a+b+c+d}.$$

Kun 0-0 ominaisuuden puuttuminen on irrelevanttia ja havainnot halutaan sivuuttaa kokonaan, käytetään samanlaisuusmittana seuraavaa:

$$\frac{a}{a+b+c}.$$

Samankaltaisuuksien ja erilaisuuksien (*matches and mismatches*) suhdetta voidaan kuvata esimerkiksi yhtälöllä

$$\frac{a}{b+c}.$$

(Johnson & Wichern, 2007, s.675.)

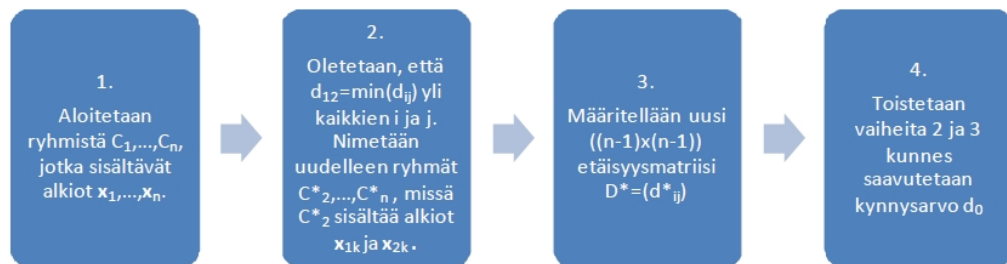
4.2.2 Hierarkinen ryhmittely

Hierarkinen ryhmittely perustuu jonoon sisäkkäisiä ryhmityksiä (Mellin, 2007).

Hierarkisissa ryhmittelymenetelmissä ryhmittely aloitetaan yhdistämällä mahdollisimman samankaltaisia yksittäisiä havaintoja (tai muuttujia) kunnes saadaan hyvä määrä ryhmiä eli klustereita (esim. kuvio 4.2). Prosessi voidaan myös suorittaa myös päinvastaiseen suuntaan siten että aloitetaan kaikista havainnoista ja jaetaan niitä mahdollisimman erilaisiin ryhmiin kunnes saadaan sopiva määrä ryhmiä. (Johnson & Wichern, 2007, s.680)

Hierarkisten menetelmien heikkous on se, että kun alkio on ryhmitelty johonkin ryhmään C , sitä ei voida enää uudelleen ryhmitellä. Mahdollisesti virheellisesti ryhmitellyt alkioit jäävät siis väärin ryhmiin. (Mardia ym., 1989, s.369)

Hierarkisen ryhmittelyn algoritmi voidaan esittää kuvan 4.2 lailla.



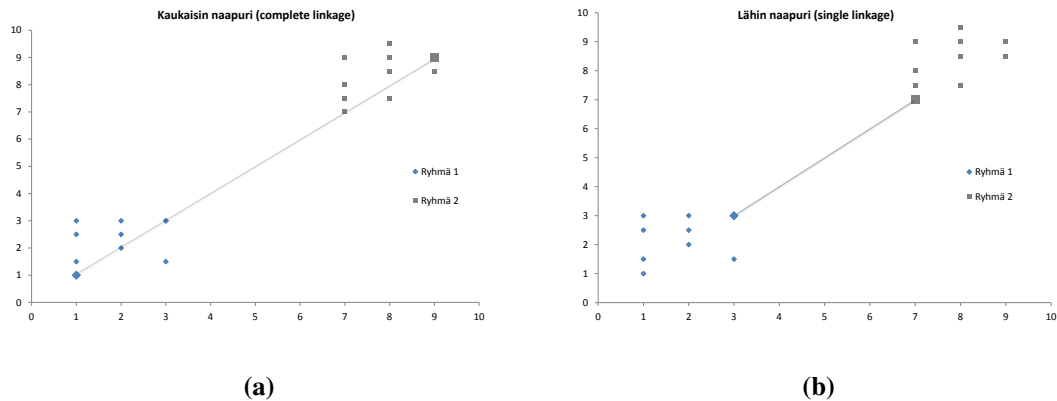
Kuvio 4.2. Hierarkisen ryhmittelyanalyysin algoritmi. Mukaillen Mardia ym. (1989, s.370)

Hierarkisia ryhmittelymenetelmiä on myös monia kuten painopistemenetelmä, lähimmän naapurin menetelmä, kaukaisimman naapurin menetelmä, keskiarvomenetelmä sekä Wardin menetelmä. Esittelen näistä lähimmän ja kaukaisimman naapurin menetelmän, sillä käytän menetelmiä aineiston analysoimiseen kohdassa 5.2.

Lähimmän ja kaukaisimman naapurin menetelmä

Kaukaisimman naapurin menetelmässä (*farthest-neighbour* tai *complete-linkage method*) etsitään nimensä mukaisesti kaukaimmaista naapuria. Kahden ryhmän välinen etäisyys saadaan määrittämällä kaikki havaintojen pareittaiset etäisyydet, joissa toinen havaintopisteistä on toisessa ryhmässä ja toinen toisessa. Näistä etäisyyksistä valitaan pisin (kuvio 4.3a). Tämä takaa, että kaikki alkioit (*item*) ovat saman maksimi etäisyyden päästä toisistaan. Kaukaisimman naapurin menetelmän etuus onkin, että se ryhmittelee havainnot usein hyvin erottuviksi ja kompakteiksi ryhmiksi (Mellin, 2007).

Lähimmän naapurin menetelmässä (*nearest-neighbour* tai *single linkage*) taas valitaan pareittaisista etäisyyksistä lyhin (kuvio 4.3b). Lähimmän naapurin menetelmä ei ole usein optimaalisin, sillä lähellä olevat havainnot kahdesta ryhmästä ryhmitellään liian herkästi yhdeksi ryhmäksi ja kun uudelleen ryhmittelymahdollisuutta ei ole, tapahtuu herkästi virheitä (Kaufman & Rousseeuw, 2005, s.47).



Kuvio 4.3. Kaukaisimman ja lähimmän naapurin menetelmä

Menetelmien käytännön tuloksia ja analyysin eroja esittelen luvussa 5.

4.2.3 Ei-hierarkinen ryhmittely

Ei-hierarkisia menetelmiä käytetään yleensä ryhmittelemään vain havaintoja. Ryhmittelyssä oletetaan, että ryhmien lukumäärästä on olemassa ennakko eli priori-tietoa, mikä voidaan katsoa menetelmän heikkoudeksi.

Menetelmän etuna on se, että sitä voi käyttää suuremmille aineistoille kuin hierarkisia menetelmiä, sillä etäisyysmatriisin ei tarvitse olla määritelty eikä tallennusta tapahdu tietokoneajon aikana (Johnson & Wichern, 2007, s.696–703).

Ryhmittely voidaan aloittaa joko enakkoryhmittelystä jonkin muuttujan suhteen, tai määrätä siemenpisteitä (*seed points*), jotka muodostavat ryhmien ytimet. Siemenpisteiden valinta tapahtuu sattumanvaraisesti, mikä heikentää menetelmän luotettavuutta (Johnson & Wichern, 2007).

Enakkoryhmittelyssä jokainen havainto kiinnitetään johonkin enakkoon valituista ryhmistä. Aloitusryhmitystä pyritään parantamaan siirtämällä havaintoja ryhmästä toiseen käyttämällä jotakin ryhmien yhtenäisyyttä kuvaavaa mittaa. Lopullinen ratkaisu saavutetaan, kun siirrot eivät enää paranna ryhmien yhtenäisyyttä. Tätä ryhmittelyä on syytä toistaa usealla eri aloitusryhmityksellä, sillä menettely on herkkä aloitusryhmityksen valinnan suhteen. Lisäksi priori-tieto ryhmien lukumäärästä on usein heikkoa, joten menettelyä tulisi toistaa myös usealle eri ryhmien lukumäärälle. (Mellin 2007, s.14.)

K-means-menetelmä

K-means on eniten käytetty ei-hierarkinen ryhmittelymenetelmä (Kaufman & Rousseeuw, 2005, s. 113). K-means-menetelmä asettaa jokaisen havainnon ryhmään, jolla on

lähin keskus ryhmiteltävän alkion kanssa (Johnson & Wichern, 2007, s.696). Keskukse-
na käytetään usein havaintojen ryhmäkohtaisten aritmeettisten keskiarvojen määräämiä
pisteitä (Mellin, 2007, s.15).

Menetelmässä on kaksi tapaa toimia. Molemmissa ensin määrätään ryhmäkeskipis-
teet ja allokoidaan jokainen havainto siihen ryhmään, jonka keskipistettä lähimpänä se
on. Tämän jälkeen kun kaikki havainnot on saatu allokoiduksi, voidaan ensimmäisen
vaihtoehdon mukaan päivittää ryhmäkeskipisteet ja suorittaa uusi allokointi. Havaintojen
allokoinnin ja ryhmäkeskipisteiden päivittämisen vuorottelua jatketaan, kunnes uusien ja
vanhojen ryhmäkeskusten etäisyydet muuttuvat vähemmän kuin etukäteen asetettu kyn-
nysarvo. Toisessa tavassa ryhmäkeskipisteet päivitetään välittömästi jokaisen havainnon
siirron jälkeen. Samalla lailla havaintojen allokoinnin ja ryhmäkeskipisteiden päivittä-
misen vuorottelua jatketaan, kunnes uusien ja vanhojen ryhmäkeskusten etäisyydet muut-
tavat vähemmän kuin etukäteen asetettu kynnsarvo. (Mellin, 2007.)

Klusterin tai ryhmän v keskus on piste p - ulotteisessa avaruudessa, ja se on saatu otta-
malla keskiarvo muuttujasta f ryhmässä v . Kun havainto i siirretään ryhmästä v ryhmään
 w , saadaan uudet keskukset seuraavasti

$$\bar{x}_f(v') = \frac{1}{n_v - 1} (n_v \bar{x}_f(v) - x_{if})$$

ja

$$\bar{x}_f(w') = \frac{1}{n_w - 1} (n_w \bar{x}_f(w) - x_{if}),$$

missä v' ja w' ovat uusia ryhmiä. (Kaufman & Rousseeuw, 2005, s.113–114.)

Menetelmän heikkoutena on, että tulokset riippuvat (suuresti) aineiston järjestyksestä
(Kaufman & Rousseeuw, 2005, s.114). Kuten muutkin mainitut ryhmittelymenetelmät,
on k -means herkkä poikkeaville havainnoille.

K-medoidi-menetelmä

K-medoidi -menetelmä on kehitetty k -means menetelmästä. Menetelmä käyttää ryhmän
edustajana medoidia keskiarvon sijasta. Menetelmä on robustimpi, sillä etäisyyden mit-
taaminen perustuu parittaisiin erilaisuuksiin (*dissimilarities*), jotka minimoidaan. Mene-
telmä kestää siis paremmin poikkeavia havaintoja ja kohinaa, eikä ryhmän keskus eli
medoidi ole riippuvainen siitä onko tieto diskreettiä vai jatkuvaa. (Kaufman & Roussee-
uw, 2005, s.117.)

Olkoon $X = x_1, x_2, \dots, x_n$ ja etäisyys kahden pisteen (x_i, x_j) välillä on $d(i, j)$ kuten
edellä. Olkoon y_i 0-1 -muuttuja niin, että se saa arvon 1 kun i ($i = 1, \dots, n$) valitaan
edustavaksi kohteeksi tai havainnoksi (*object*). Olkoon z_{ij} myös 0-1 -muuttuja, joka saa
arvon 1 kun havainto j on asetettu ryhmään, jossa i on edustava havainto ja siis medoidi.
Etäisyys minimoidaan seuraavan yhtälön mukaisesti.

$$\min \sum_{i=1}^n \sum_{j=1}^n d(i, j) z_{ij},$$

missä

$$\sum_{i=1}^n z_{ij} = 1 \text{ kun } j = 1, 2, \dots, n, z_{ij} \leq y_i \text{ kun } i, j = 1, 2, \dots, n,$$

$$\sum_{j=1}^n y_j = k, k \text{ on ryhmien lukumäärä ja } y_i, z_{ij} \in 0, 1.$$

(Kaufman & Rousseeuw, 2005, s.108–109.)

Menetelmän sovellus on Kaufman & Rousseeuw (2005) kehittämä PAM-algoritmi (*Partitioning Around Medoids*), joka koostuu kahdesta vaiheesta (BUILD eli rakenna ja SWAP eli vaihda). Ensimmäisessä askeleessa valitaan satunnaisesti k kappaletta kaikista n medoideiksi ja tämän jälkeen jokaiselle havainnolle n etsitään lähin medoidi ja siis ryhmä. Muuttujan havaintoarvot myös normitetaan. Toisessa vaiheessa on etsitään ryhmille parhaat uudet edustajat. (Kaufman & Rousseeuw, 2005, s.102.)

Sovellus tuottaa parempia ryhmiä kun poikkeavat havainnot eivät sotke ryhmitystä eikä se riipu havaintojen järjestyksestä (Kaufman & Rousseeuw, 2005, s.117). Ryhmien onnistuneisuutta voidaan arvioida graafisesti siluettikuvien avulla (Kaufman & Rousseeuw, 2005, s.68). Kuvissa esitetään silutettiarvot ryhmittäin havaintonumeroita vasten ja negatiiviset arvot kertovat huonosta ryhmittelystä (Vehkalahti, 2008, s.168)

5 Kyselyn tulosten analysointi monimuuttujamenetelmillä

Analysoin aineistoani esittelemilläni monimuuttujamenetelmillä eli faktorianalyysillä sekä ryhmittelyanalyysillä.

Alkuperäisenä tarkoituksena oli muodostaa faktorit kuvaamistani ulottuvuuksista 3.2 ja verrata faktoripisteitä sekä niiden erilaisuuksia maksavien ja ilmaiseksi opiskelevien välillä. Nyt kun aineistosta ei ole mahdollisuutta vertailua tehdä käytän faktoripisteitä ryhmittelemään opiskelijajoukkoa, jotta saataisiin selville minkälaiset opiskelijat ovat maksumyönteisiä tai -kielteisiä. Ryhmittelyanalyysillä olisi voinut tarkastella maksavien ja ilmaiseksi opiskelevien sijoittumista erilaisiin ryhmiin faktoripisteiden mukaan. Ryhmien eroja olisi voinut tulkita vielä esimerkiksi erotteluanalyysillä.

Kuten aineiston tarkasteluosiossa 3.4 totesin, ei keräämäni aineisto välttämättä edusta perusjoukkoa, joten tässä luvussa esitettyihin tuloksiin tulee suhtautua varauksella. Lisäksi aineisto ei sovi täydellisesti analyysimenetelmien käyttöön muuttujien hyvin viinon jakaumien vuoksi. Empiiriset analyysit ovat siis tässä työssä esimerkinomaisia.

5.1 Faktorianalyysin tulokset

Mittausmallin avulla (kuvio 3.2, s.25) hahmottaa helposti tutkijan ajatuksen mahdollisesta faktorirakenteesta. Mittausmallissa katkoviihvin piirretyt ympyrät edustavat ennalta ajattelemiani faktoreita ja neliöt muuttujia eli todellisia kysymyksiä. Lähdin siis tekemään faktorianalyysia konfirmatorisella eli teoriaan pohjautuvalla tavalla, mutta pysähdyin tulkitsemaan tuloksia ottaen eksploratiivisen näkökulman aiheeseen teoriapohjan ollessa heikko.

Määrittelin faktoreiden luvuksi neljä mittausmallin mukaan, muodostin muuttujien välisen korrelaatiomatriisin, jonka faktoroin suurimman uskottavuuden menetelmällä. Ohjelmakoodit löytyvät liitteestä G. Tarkastelin tulosteita ja havaitsin mm. *Scree-plot*:in avulla, että analyysin mukaan aineistosta löytyisi seitsemän faktoria. Tämä tuntui kuitenkin epäselvältä ratkaisulta ja lähemmin korrelaatiomatriisia tutkittuani huomasin viisi selkeämpää ryhmittymää ja tein analyysit uudelleen käyttäen viittä faktoria. Selkeämmän ratkaisun saamiseksi rotatoin saamani tuloksen Varimax–menetelmällä. Tulokset löytyvät taulukosta 5.1.

Taulukon 5.1 toiseksi viimeinen sarake (*Sumsqr*) kertoo kysymysten **kommunaliteetit** eli kuinka hyvin faktori selittää kyseisen muuttujan vaihtelua (KvantiMOTV). Kommunaliteetti kertoo vaihtelun suhteellisesta osuudesta eli mitä lähempänä luku on yhtä, sitä paremmin osio toimii mittausmallissa (Vehkalahti, 2008, s.99). Kommunaliteetti eli systemaattinen vaihtelun osuus saadaan laskemalla faktorilatausten neliösumma seuraa-

valla kaavalla:

$$h_i^2 = a_{i1}^2 + a_{i2}^2 + \dots + a_{ip}^2, i = 1, 2, \dots, p,$$

jossa a tarkoittaa faktorilatauksia. (Mustonen, 1995, s.76.)

Taulukko 5.1. Faktorilataukset

Name	F1	F2	F3	F4	F5	Sumsqr	Label
im_A_7	0,59	0,09	0,20	0,05	0,10	0,41	Quality of student facilities
im_A_6	0,58	0,06	0,31	0,08	0,12	0,46	Quality of research
im_FI_11	0,54	0,01	0,25	0,11	0,16	0,39	Reputation of Finnish education
im_FI_12	0,53	0,05	0,26	0,23	0,01	0,41	Reputation of Finnish industry
im_A_8	0,52	0,05	0,42	0,04	0,07	0,46	Quality of teaching
im_A_4	0,51	0,05	0,46	-0,03	0,10	0,48	Good reputation of my study program
im_A_1	0,47	0,14	0,02	0,13	0,04	0,26	Academic freedom/independence
im_FI_10	0,46	0,07	0,07	0,11	0,14	0,26	Pursuing a university degree
im_A_11	0,45	0,14	0,19	0,18	-0,05	0,29	Scholarship
im_A_12	0,43	0,10	0,16	0,34	-0,04	0,34	Social life
im_FI_13	0,36	0,29	0,14	0,29	0,06	0,33	Visa Processing system
im_FI_3	0,32	0,00	0,08	0,08	0,03	0,12	Experiencing a foreign country
im_VM_3	0,15	0,77	0,02	0,07	-0,08	0,62	My family took a loan to cover my study costs
im_VM_5	0,08	0,71	-0,01	0,14	-0,04	0,53	... sold property to cover my study costs
im_VM_7	0,16	0,61	0,01	0,01	0,15	0,41	... works extra hard to cover my expenses
im_VM_1	0,28	0,43	0,01	0,01	0,19	0,30	My master's is big investment for my family
im_VM_8	-0,02	0,42	0,04	-0,04	0,21	0,23	I had/have to work extra hard to cover my expenses
im_VM_6	0,00	0,42	0,03	0,09	0,03	0,19	I sold some of my property to cover my study costs
im_VM_4	0,01	0,42	0,01	-0,01	0,05	0,18	I took a personal loan to cover my study costs
im_VM_2	0,12	0,33	0,05	0,04	0,24	0,19	My master's is big investment for myself
im_A_3	0,19	0,07	0,71	0,06	0,16	0,58	Good reputation of the university
im_A_10	0,19	0,01	0,66	0,17	0,07	0,50	Reputation of a qualification
im_A_13	0,26	0,09	0,48	0,15	0,11	0,34	University's ranking position
im_A_2	0,12	0,13	0,47	0,37	-0,05	0,39	Employment chances after studies
im_A_9	0,16	0,02	0,38	0,07	-0,05	0,18	Recommendation
im_FI_8	0,15	0,13	0,05	0,68	0,06	0,50	Possible migration/citizenship
im_FI_6	0,19	-0,05	0,10	0,55	0,16	0,38	High standard of living in Finland
im_FI_7	0,12	0,04	0,13	0,49	-0,04	0,27	Learning Finnish/Swedish
im_FI_5	0,15	0,17	-0,05	0,47	0,08	0,28	Hardship in home country
im_FI_4	-0,25	0,02	0,07	0,41	-0,18	0,27	Family ties (incl. boy/girlfriend)
im_FI_9	0,21	-0,05	0,05	0,37	-0,06	0,19	Prior knowledge of Finland
im_FI_1	0,26	0,14	0,21	0,29	0,08	0,22	Career development/work opportunities
im_V_7	0,06	0,16	0,01	0,07	0,61	0,41	Tuition fees prevent poor students from pursuing HE
im_V_1	0,07	0,14	0,04	0,05	0,47	0,25	Free education is a basic right
im_V_4	0,01	0,09	0,04	-0,11	0,45	0,22	Scholarships should be given to poor students
im_FI_2	0,25	0,21	-0,04	0,06	0,43	0,30	Cost of study
im_A_5	0,16	-0,01	0,10	0,08	0,32	0,15	Possibility of studying in English
im_V_2	0,03	0,07	-0,06	0,04	-0,42	0,19	High tuition fees guarantee a high quality of education
im_V_5	0,11	-0,14	0,24	-0,10	0,09	0,11	Scholarships should be given to the students with best
im_V_8	0,10	-0,11	-0,02	-0,06	0,02	0,03	Tuition fees should be used to improve the quality of teaching, even if non-paying students also benefit from it
Sumsqr	3,57	2,67	2,44	2,23	1,66	12,57	

Taulukkoa 5.1 tarkastelemalla huomataan, että muuttujien kommunaliteetit eivät ole kor-

keita ja suurin kommunaliteetti 0,62 on muuttujalla *im_VM_3* ja pienin kommunaliteetti 0,03 muuttujalla *im_V_8*. Tarkasteltuani yksittäisiä muuttujia, joiden kommunaliteetit olivat hyvin alhaisia, huomasin, että vastaukset olivat keskittyneet tiettyihin vaihtoehtoihin ja niiden hajonta oli hyvin pientä. Kysymykseen *im_V_8* ”*Tuition fees should be used to improve the quality of teaching, even if non-paying students also benefit from it*” -kysymykseen oli 53 % vastannut ”*Totally agree*” (liite E) . Kun faktorille lausautuvat muuttujien hajonta ei ole suuri, ei hajonnan selitysosuuskään voi olla korkea.

Faktorien voimakkuuksilla tutkitaan faktorimallin toimivuutta. **Voimakkuudet** näkyvät taulukon 5.1 viimeisellä rivillä ja niiden summa eli yhteisvaihtelu on myös kommunaliteettien summa. Voimakkuudet kertovat kuinka paljon faktorit selittävät muuttujien hajonnasta; mitä suurempi voimakkuus sitä paremmin faktori selittää muuttujien hajontaa. Paremmin vertailtava suhteellinen selitysosuus saadaan jakamalla voimakkuus havaittujen muuttujien lukumäärällä, jolloin luku saa arvoja välillä [0,1]. Saatu luku kertoo siis kuinka paljon faktori selittää kaikkien malliin otettujen muuttujien hajonnasta. Kun kaikkien faktoreiden selitysosuudet lasketaan yhteen, saadaan tietää kuinka paljon faktorit yhteensä voivat selittää malliin valittujen muuttujien hajonnasta. (Vehkalahti, 2008, s.99.)

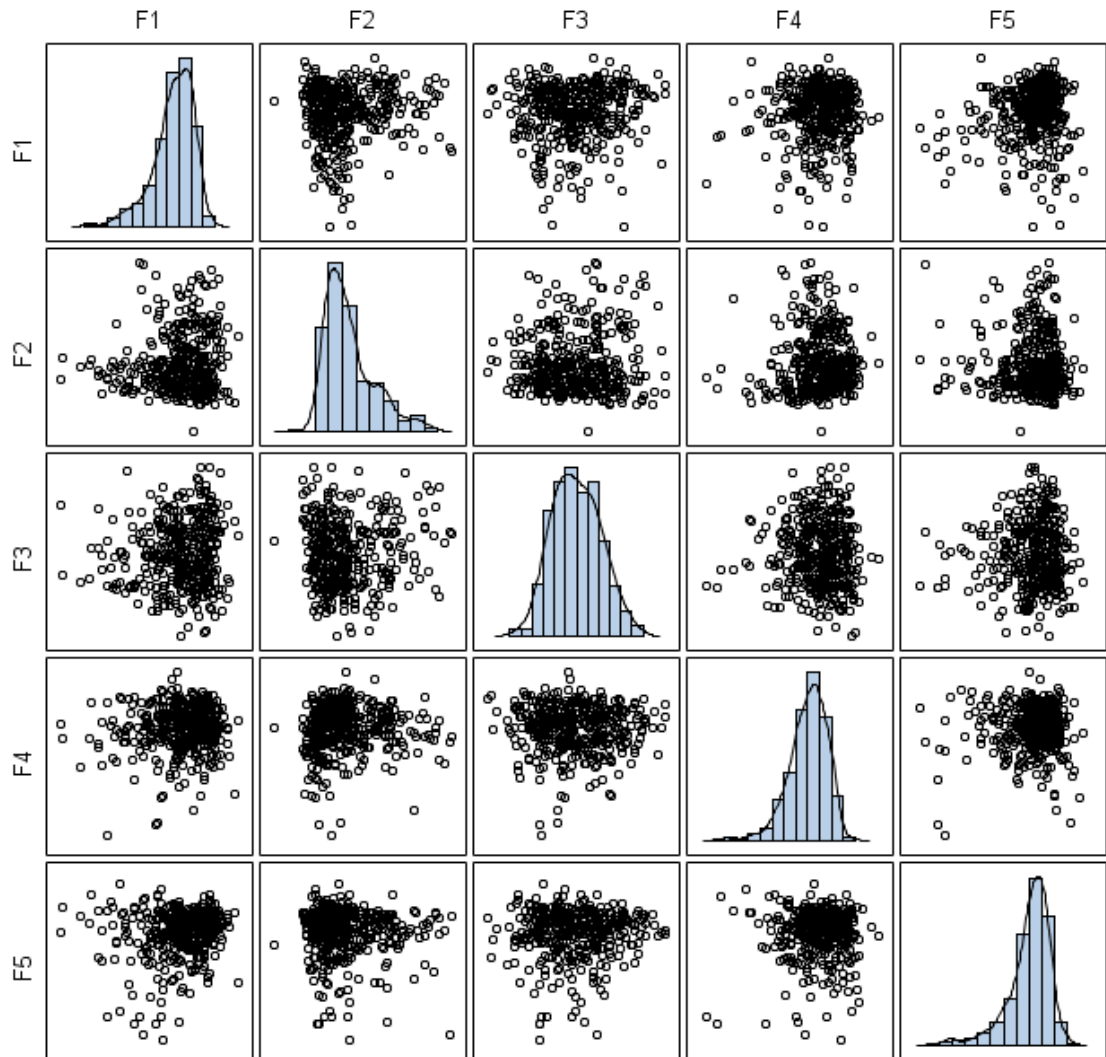
Muodostamieni faktorien voimakkuudet näkyvät myös taulukosta 5.2, johon olen laskenut myös suhteelliset selitysosuudet. Jokainen faktori yksinään selittää alle 10 % muuttujien vaihtelusta. Yhteensä ne selittävät siis 32 % kokonaisvaihtelusta, jolloin jopa 68 % jää mittausvirheille. Yleensä kyselytutkimuksissa selitysosuus on 50 %, joten analyysin tulos ei ole niin huono miltä se vaikuttaa (henkilökohtainen viestintä, Vehkalahti, 27.3.2012). Faktoreita lisättäessä myös voimakkuudet ja niiden summat kasvavat, kuitenkin faktoreiden runsas määrä ei ole tutkimuksen kannalta tarkoituksenmukaista. Imputoitukaan aineisto ei ole siis täydellinen faktorianalyysin tekoon ja tulkinta jää kevyeksi, sillä faktorit eivät selitä kovin paljoa muuttujien vaihtelusta.

Taulukko 5.2. Faktoripistemuuttujien nimet, kuvaukset ja selitysosuudet

Tunnus	Nimi	Kuvaus	Voimakkuudet	Selitysosuus	Selitysosuus faktoreista
F1	Houkutus	Koulutuspaikan houkuttelevuus	3,57	9 %	28 %
F2	Investointi	Koulutus investointina	2,67	7 %	21 %
F3	Arvo	Koulutuksen arvo	2,44	6 %	19 %
F4	Suomeen	Suomi kohteena	2,23	6 %	18 %
F5	Maksuttomuus	Koulutuksen maksuttomuus	1,66	4 %	13 %
			12,57	32 %	100 %

5.1.1 Faktoripisteet

Faktoripisteet (*factor scores*) ovat faktoreita kuvaavia havaintokohtaisia arvoja, joita voi käyttää mm. ryhmittely- tai erotteluanalyysin tekoon tai regressiomalli selittäjinä (Vehkalahti, 2008, s.109). Faktoripisteiden käyttö edellyttää, että faktorit pysytään nimeämään eli että ne ovat hyviä ja selkeästi tulkittavia uusia muuttujia (Vehkalahti, 2008, s.109). Faktoreiden nimet ja kuvaukset on nähtävissä taulukosta 5.2. Nimet pyrkivät edustamaan muodostamieni faktoreiden tulkintaa. On kuitenkin pidettävä mielessä, että nämä tulokset ovat enemmänkin esimerkinomaisia. Faktoripisteet ovat menetelmän ole-



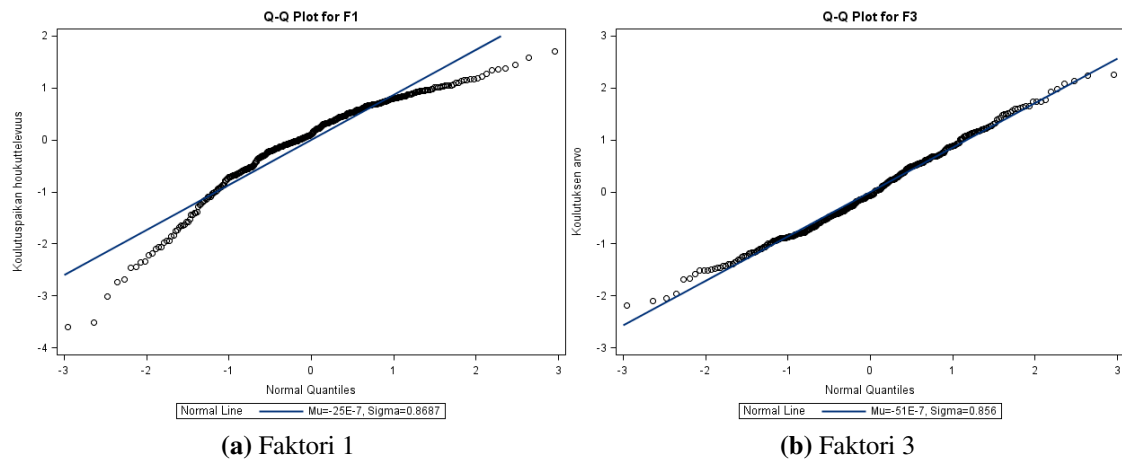
Kuvio 5.1. Faktoripisteiden hajontakuvat ja histogrammit

tusten mukaan normaalisti jakautuneita eivätkä ne korreloi keskenään, kun rotaatiomenetelmäksi on valittu ortogonaalinen rotaatio. Faktoripisteet on skaalattu niin, että niiden keskiarvot ovat nollia. (Mustonen, 1995.)

Käyttämäni SURVO-ohjelma tuottaa faktorien kerroinmatriisin K (kuten yhätlössä 4.3) sukrolla `/FCOEFF` käyttäen regressiomenetelmää ja faktoripisteet jokaiselle havainnolle `LINCO`-komennolla (liite G) (Mustonen, 1995, s.91). Muodostamani faktoripisteet eivät kaikki kuitenkaan ole normaalisti jakautuneita, mikä nähdään silmämääräisesti kuvion 5.1 histogrammeista. Myös kvantiilikuvioita 5.2 tarkastellessa nähdään, että vain faktori 3 on normaalisti jakautunut (pisteet sijaisevat kuvaajassa suoralla). Ainoastaan faktori 3 eli muuttuja *Koulutuksen arvo* on jakautunut melko normaalisti ja faktoripisteiden hajontakuvioista nähdään, että ainoa hyvin joukkoa erotteleva muuttuja on faktori 3. Tämä johtunee siitä, että myös alkuperäiset muuttujien hajonnat olivat hyvin vinoja (ks. liite E).

Tässä työssä käytän faktoripisteitä ryhmittelyanalyysin tekoon, sillä ne ovat iän lisäksi ainoita jatkuvia muuttujia aineistossa ja ne ovat valmiiksi skaalattu samalle asteikolle. Kuten kuvasta 5.1 nähdään, faktoripisteiden jakaumat ovat kuitenkin selvästi

vinoja. Havainnot keskittyvät paljon ääripäihin ja vain faktori F3 olisi teoriassa sopiva ryhmittelyanalyysin tekoon.



Kuvio 5.2. Kvantiilikuviot faktoreille 1 ja 3

5.2 Ryhmittelyanalyysi faktoripisteillä

Käytän opiskelijoiden esimerkinomaiseen ryhmittelyyn ensin hierarkista-menetelmää osa-aineistolle, jotta saan kuvan tai hienostuneet arvauksen ryhmien lukumäärästä (Malhotra & Birks, 2007, s.680). Lopulta ryhmittelen koko aineiston opiskelijat *k-means* menetelmällä.

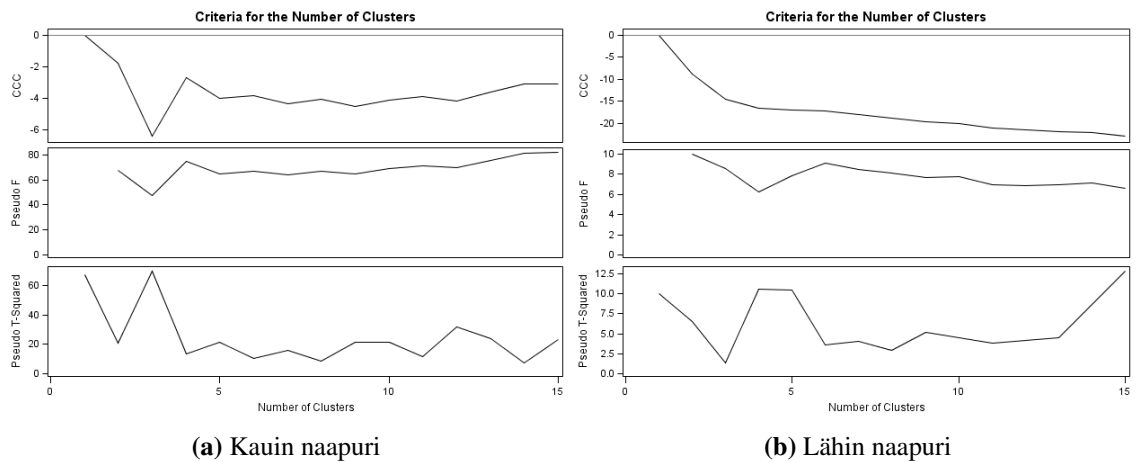
5.2.1 Hierarkinen ryhmittely

Hierarkinen ryhmittelyanalyysi sopii yleensä paremmin pienemmille aineistoille, joten esimerkin selkeyden vuoksi tein analyysin vain aineiston osajoukolle. Päätin ottaa aineistoon kansallisuudeltaan eurooppalaiset, sillä heitä oli sopiva määrä aineistossa, 88 kappaletta.

Ryhmittelymuuttujiksi valitsin faktorit 3 ja 5. Faktorin 3 valitsin sen teoreettisen hyvyyden vuoksi ja faktorin 5 (*Koulutuksen maksuttomuus*) sen sisällön vuoksi. Tein ryhmittelyanalyysin kahdella erilaisella hierarkisen ryhmittelyn menetelmällä, kauimman ja lähimmän naapurin menetelmällä, jotka esittelin osiossa 4. Ryhmittelyanalyysin SAS-koodi löytyy liitteistä G.

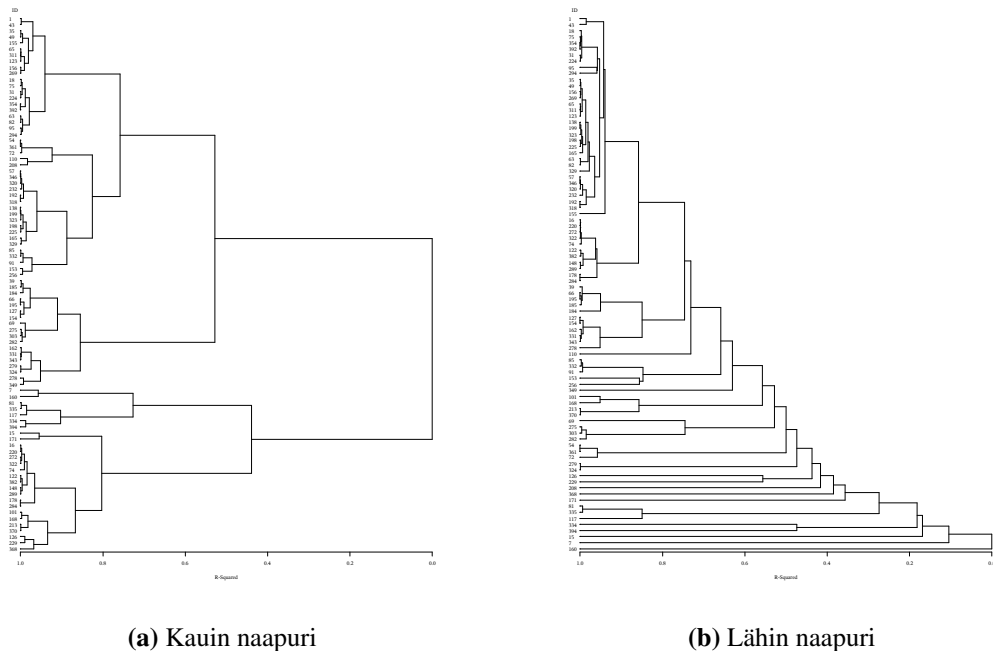
Hierarkisessa ryhmittelyssä klustereiden eli **ryhmien määrää** voi arvioida puukuvan, numeeristen tulosteiden ja niistä piirrettyjen graafien avulla. SAS-ohjelma tulostaa `proc cluster` – proseduurissa numeeriset tulosteet ja kolmen muuttujan kuvaajan, joista ryhmien lukumäärää voi arvioida. Kuvion 5.3 ensimmäisiä kuvaajia CCC (*cubic clustering criterion*) eli kuutioitu ryhmittelykriteeriä voidaan tulkita niin, että yli 2 tai 3 arvon saavat huiput ovat merkinä hyvistä klusterimääristä ja huiput 0 ja 2 välissä merkitsevät mahdollisia klustereita. Hyvin isot negatiiviset luvut voivat olla merkinä poikkeavista havainnoista. Toisen kuvaajan, *Pseudo F statistics*, hyvin isot arvot merkitsevät hyviä klusterimääriä. Kolmatta kuvaajaa *Pseudo t^2 statistics* tulee lukea oikealta

vasemmalle kunnes vastaan tulee ensimmäinen selvästi isompi arvo, joka voidaan tulkita hyväksi ryhmämääräksi. (Cary, 2011.)



Kuvio 5.3. Kolme kuvaajaa tulosteista

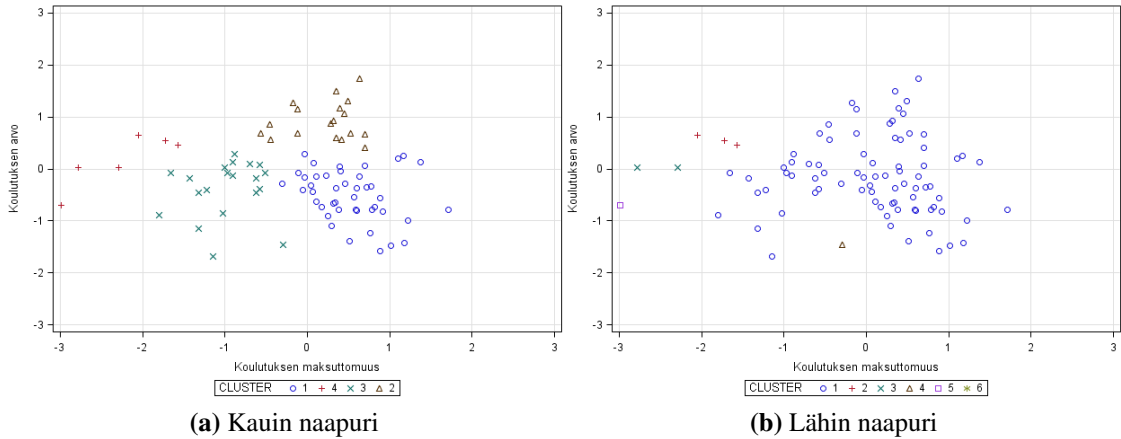
Molempien menetelmien *CCC* – kuvaajat (kuvio 5.3) ovat negatiivisen puolella, mikä ilmeisesti kertoo poikkeavista havainnoista tai faktorin 5 vinoudesta. *Pseudo F* – kuvaajan mukaan kauimman naapurin menetelmällä hyvä ryhmien lukumäärä olisi neljä ja lähimmän naapurin menetelmällä kuusi. Toisaalta kumpikaan kuvaaja ei saa ”hyvin isoja” arvoja. *Pseudo t^2* – kuvaajan mukaan ryhmien lukumäärät olisivat kuten edellä.



Kuvio 5.4. Hierarkisen ryhmittelyn tulokset puukuviaina

Ryhmiä selityksastetta eli R^2 voi arvioida numeerisesta tulosteesta tai puukuviosta, jossa R^2 löytyy vaaka-akselilta. Mitä suurempi R^2 arvo sitä enemmän ryhmien määrä

selittää vaihtelusta. Kauimman naapurin menetelmällä (kuvio 5.4a) ryhmien lukumääräksi vahvistuisi neljä, sillä kuvaajan haarautuessa neljään osaan R^2 on 0,6 ja 0,8 välissä (numeerisesta tulosteesta tarkastettuna 0,73) mikä tarkoittaa sitä että neljä ryhmää selittäisi 73 % ryhmien vaihtelusta. Kuvioista 5.4b nähdään, että lähimmän naapurin menetel-

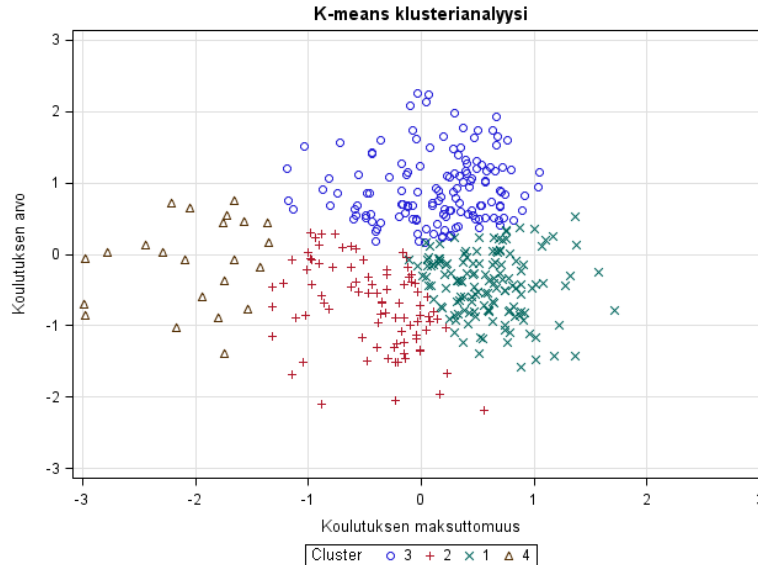


Kuvio 5.5. Hierarkisen ryhmittelyn tulokset hajontakuvioiden muodossa

mä ei ole onnistuneesti pystynyt ryhmittelemään havaintoja vaan ne ovat kaikki yhdessä kasassa. Tämä oli odotettavaa, sillä, kuten jo totesin luvussa 4, menetelmä ei ole optimaalinen ryhmittelyyn. Lähellä olevat havainnot kahdesta ryhmästä ryhmitellään liian herkästi yhdeksi ryhmäksi ja virheitä tapahtuu herkästi. Kuviossa 5.5 on esitetty lähimmän ja kauimman naapurin menetelmällä saadut ryhmät eri symboleilla kahden ryhmittelymuuttujan hajontakuvaan. Nähdään, että kauimman naapurin menetelmällä (kuvio 5.5a) ryhmittely on onnistunut melko hyvin, sillä ryhmät ovat helposti havaittavissa, eri symboleita on suhteellisen tasainen määrä ja ne ovat melko hyvin erillään. Lähimmän naapurin menetelmällä (kuvio 5.5b) ryhmittely on onnistunut kohtalaisen huonosti, sillä melkein kaikki havainnot kuuluvat ryhmään 1.

5.2.2 Ei-hierarkinen ryhmittely ja ryhmien tulkinta

Ei-hierarkista ryhmittelyä ja useasti *k-means* -menetelmää sovelletaan yleensä isommille aineistoille, joten valitsin analyysiini koko aineiston. Menetelmä vaatii ennakkotiedon ryhmien lukumäärästä ja sitä käytetäänkin usein hierarkisen klusteroinnin luoman ryhmittelyn pohjalta, isommalle aineistolle, paremman laskennallisen kapasiteetin vuoksi. Käytin analyysissä siis hierarkisella, kauimman naapurin menetelmällä saamaani ryhmien lukua 4. Tein ryhmittelyn samojen muuttujien, faktoreiden 3 ja 5, suhteen kuten edellä. En tule siis analysoimaan ryhmittelyn hyvyttä *k-means* -menetelmällä, vaan käytän menetelmää vain ryhmittelyn tekemiseen koko aineistolle. Ryhmittelyn onnistuneisuutta sekä ryhmien ominaisuuksia voi arvioida SAS-ohjelman `FASTCLUS`-proseduurin tulosteesta, mistä löytyy siemenlukuina käytetyt klusterikeskukset, ryhmien koot sekä erottelumuuttujien keskiarvot ja hajonnat ryhmittäin. Ryhmien ominaisuuksien tulkinta lähtee ryhmittelymuuttujina käytettyjen muuttujien ryhmäkeskiarvojen tulkinnasta. Lisäksi voidaan tutkia ryhmittelyanalyysin ulkopuolelle jätettyjä muuttujia kuten demografisia tekijöitä.



Kuvio 5.6. Ei-hierarkisen ryhmittelyn tulos hajontakuviona

Ryhmien kuvaus

Kuviossa 5.6 on esitetty hajontakuva erottelumuuttujien suhteen *k-means* ryhmittelyn laskemilla ryhmillä. Taulukosta 5.3 nähdään ryhmäkohtaiset keskiarvot sekä hajonnat faktoripistemuuttujille. Keskiarvot ryhmittäin ovat piirrettynä myös kuvioihin 5.7a ja 5.7b.

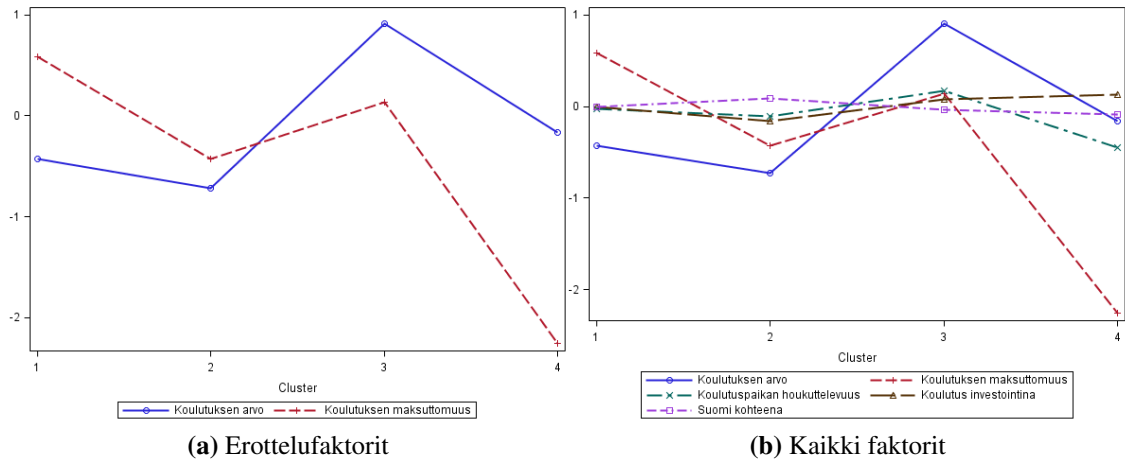
Taulukko 5.3. K-means ryhmittelyn ryhmäkeskiarvot ja hajonnat

Ryhmä	Frekvenssi	Keskiarvo		Keskihajonta	
		F5	F3	F5	F3
1	139	0,59	-0,43	0,35	0,45
2	88	-0,43	-0,72	0,43	0,59
3	140	0,14	0,91	0,48	0,48
4	28	-2,26	-0,16	0,73	0,63

Hajontakuvan 5.6 mukaan ryhmä 1 (symboli x) pitää koulutuksen maksuttomuutta tärkeänä ja antaa saamalleen koulutukselle yhtä paljon arvoa kuin ryhmät 2 ja 4. Ryhmä 2 (symboli +) ei pidä koulutuksen maksuttomuutta niin tärkeänä kuin ryhmä 1, ja ryhmä 4 (symboli \triangle) kahta edellistä vielä vähemmän. Ryhmä 3 (symboli \circ) arvostaa saamaansa koulutustaan enemmän kuin muut ryhmät mutta jakaantuu enemmän mieliteissään koulutuksen maksuttomuudesta. Ryhmässä 4 hajonta molempien muuttujien suhteen oli korkeampaa kuin muissa ryhmissä kuten taulukosta 5.3 nähdään.

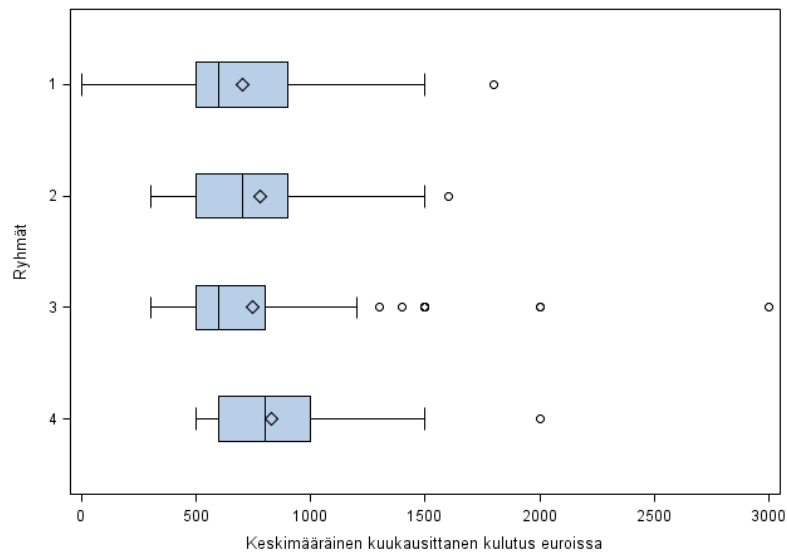
Kuvasta 5.7 nähdään, että ainoastaan ryhmä 4 suhtautuu keskimäärin reilusti muita ryhmiä negatiivisemmin koulutuksen maksuttomuuteen.

Kuvioilla 5.8 sekä liitteistä löytyvillä kuviolla F.1–F.4 olen pyrkinyt kuvaamaan lisää ryhmien ominaisuuksia aineistosta löytyvien taustamuuttujien avulla. Kuviossa 5.8 on esitetty opiskelijoiden ilmoittamat kuukausittaiset menot laatikkokuvan avulla. Kuvioon F.1 (liitteet) on piirretty opiskelijoiden isien koulutustasot osuuksina. Näemme, että useimmilla opiskelijoilla on korkeakoulutetut vanhemmat. Kuvioista F.2 näemme mistä kouluista ryhmien opiskelijat tulevat ja kuvioista F.3 näemme minkä ikäisiä he



Kuvio 5.7. Faktoripisteiden keskiarvot ryhmittäin, k-mean ryhmittely tulos koko aineistolle

ovat. Kuviossa F.4 on käytetty hyväksi Maailmanpankin aineistoa, jossa maat on luokiteltu viiteen kategoriaan elintasonsa mukaan. Opiskelijoiden ilmoittamat kansallisuudet on luokiteltu vastaamaan Maailmanpankin luokittelua ja niiden osuudet ovat nähtävissä kuvioista.



Ryhmän 1 opiskelijoiden isät ovat kandidaatteja sekä maistereita. Opiskelijoista 90 % on alle 30-vuotiaita. Kun katsotaan kuviota F.2 jopa 25 % ei ole paljastanut koulutusohjelmaansa tai se on Erasmus Mundus-ohjelma. Ryhmässä on myös opiskelijoita korkean tulotason ei-OECD-maista kuten Singaporesta. Ryhmään 1 kuuluvat käyttävät keskimäärin vähiten rahaa kuukaudessa (kuvio 5.8), mutta hajontaa joukossa on paljon, sillä minimi ja maksimi ovat kaukana toisistaan.

Ryhmässä 2 isän yleisin koulutustaso on myös kandidaatti. Opiskelijat ovat enimmäkseen 25 – 29 -vuotiaita ja opiskelevat taide- ja insinöörialoja. Tässä ryhmässä on suhteellisesti vähiten kaikkein alimman tulotason maista kuten Bangladeshista sekä Ne-

palista. Kuitenkaan opiskelijoiden ilmoittamat kulut eivät poikkea merkittävästi muista ryhmistä.

Ryhmässä 3 on suurin suhteellinen osuus opiskelijoista *Upper middle income* -maista kuten Kiinasta ja Venäjältä. Ryhmässä on myös suhteellisesti eniten opiskelijoita kauppa- sekä perustieteiden korkeakouluista ja nuorimpia eli 21 – 24 -vuotiaita. Isän koulutustaso mukailee samaa rakennetta kuin muissakin ryhmissä, mutta tässä ryhmässä on suhteellisesti eniten vain peruskoulun suorittaneita.

Ryhmässä 4 opiskelijoiden isät ovat korkeammin koulutettuja kuin muissa ryhmissä, ryhmässä näyttää olevan eniten tohtoreita sekä maistereita. Lisäksi ryhmässä 4 on eniten taidealan opiskelijoita ja heistä 40 % on alle 25 vuotiaita. Suurin osa (80 %) ryhmän 4 opiskelijoista tulee korkean elintason maista. Ryhmä käyttää keskimäärin eniten rahaa elämiseen kuukausittain. Ryhmä 4 on pienin lukumäärältään, joten poikkeamat voivat vääristävät tuloksia. Toisaalta ryhmä sopisi hyvin uskomuksiin siitä, että tällainen elitistinen ryhmä kannattaa lukukausimaksuja.

Ryhmien nimeäminen on tässä kontekstissa haastavaa eri kokoisten ryhmien sekä epämääräisten taustamuuttujien vuoksi. Taulukosta 5.4 nähdään ryhmien ominaisuuksien yhteenveto ja keksimäni nimet ryhmille. Näihinkin tulee suhtautua esimerkinomaisesti.

Taulukko 5.4. Ryhmien kuvaukset

Muuttuja	Ryhmä1	Ryhmä2	Ryhmä3	Ryhmä4
Rahankäyttö/kk keskiarvo	703 €	781€	748 €	829 €
Isän koulutustaso	Kandidaatti- ja maisteri	Kandidaatti	Peruskoulua eniten	Tohtoreita
Ikä	Alle 30	25-29	21-24	21-24
Kotimaan tulotaso	muut kuin OECD	vähiten alinta luokkaa	Kiina ja Venäjä	OECD
Opiskeluala	Erasmus Mundus	taide- ja insinöörialat	kauppa- ja perustieteet	taide
Koulutuksen arvo	Ei niin tärkeää	Ei niin tärkeää	Tärkeintä	Välinpitämätön
Koulutuksen maksuttomuus	Tärkeintä	Välinpitämätön	Välinpitämätön	Ei yhtään tärkeää
Ryhmän nimi	”Vähävaraiset”	”Tavikset”	”Pyrkyrit”	”Snobit”

6 Johtopäätökset

Lukukausimaksujen esittämisellä tai maksujen korotuksella on vaikutusta etenkin alemmista sosiaalisista luokista tulevien halukkuuteen osallistua korkeakoulutukseen (Chapman & Ryan, 2005). Ylioppilaskunnat Suomessa ovat pelänneet, että 2011–2014 käynnissä oleva lukukausimaksukokeilu EU- ja ETA-maiden ulkopuolisille opiskelijoille karkottaa lahjakkaat, mutta köyhemmät opiskelijat kokonaan, jolloin opiskelijamassa homogenisoituu niin, että jäljellä ovat vain ne joilla on varaa maksaa korkeiden elinkustannusten lisäksi lukukausimaksut (Sauli, 2011).

Lukukausimaksukokeilun vaikutuksia opiskelija-aineeseen pyrittiin selvittämään Aalto-yliopiston ulkomaalaisille maisteriopiskelijoille teetetyn kyselytutkimuksen avulla. Tarkoituksena oli verrata 2010 ja 2011 aloittaneita opiskelijoita toisiinsa. Kuitenkin yllättävä pieni maksavien opiskelijoiden määrä esti vertailun teon ja tutkimus muutettiin pilotiksi sisältäen kysymyksenasettelun testauksen ja opiskelija-aineen kuvauksen mm. ryhmittelyanalyysin avulla.

Kyselytutkimuksen teossa on monta liikkuvaa osaa, joita pitäisi samanaikaisesti osata hallita (Vehkalahti, 2008, s.11). Internetissä toimeenpantavan kyselytutkimuksen suurimmat ongelmakohdat ovat kohderyhmän yhteystietojen tarkkuus, vastaajien rekrytointi sekä lomakkeen onnistuminen niin kysymysten kuin teknisten ulottuvuuksien muodossa. Tässä tutkielmassa kiinnitettiin etenkin huomiota vastaajien rekrytointiin ja vastauskadon ehkäisyyn. Esimerkiksi ennakkotiedonannolla sekä muistuttelulla pyrittiin nostamaan vastausprosenttia ja kyselyn keskeyttäneiden määrää koitettiin alentaa mm. pakollisten kysymysten vähyydellä.

Opiskelijoilta kysyttiin sosioekonomiseen taustaan liittyvien kysymysten lisäksi opiskelijoiden motivaation lähteitä hakeutua Suomeen sekä Aalto-yliopistoon, ja heiltä mitattiin asenteita koulutuksen maksullisuutta kohtaa. Kysely lähetettiin kaikille kohderyhmän 868 opiskelijalle, joten mitään otanta-asetelmaa ei käytetty. Kyselyyn vastasi yhteensä 395 opiskelijaa. Kerätty aineisto oli rekisteritarkastelujen perusteella melko edustava otos perusjoukosta, sillä ainakin aineiston ikä- ja kansallisuusjakaumat vastasivat rekisteriä suurin piirtein. Kerätyn aineiston perusteella saatiin yleiskuva opiskelijoiden asenteista sekä sosioekonomisista taustoista.

Kyselytutkimuksen vastausprosentti oli hyvä (n. 46 %), mikä johtunee opiskelijoita kiinnostavan aiheen lisäksi, siitä että tutkimukseen osallistumisen porkkanana oli 100 euron lahjakortin arvonta. Kysely suunniteltiin ja toteutettiin noudattaen kirjallisuudessa esitettyjä ohjeita ja lisäksi kyselylomakkeen kysymyksiin vastattiin tasaisesti niin, että lomaketta ei oltu täytetty esimerkiksi vain alkupäästä. Voidaan siis olettaa, että myös kyselyn toteutus ja lomake olivat vähintäänkin melko onnistuneita. Valitettavasti keskeyttäneistä ei ollut tilaisuutta saada informaatiota, joten vastauskatoanalyysi sekä lomakkeen hyvyden perusteellinen arviointi jäivät tyngäksi. Kyselyn viimeisen sivun kommenttilaatikkoon jätettiin 35 kommenttia koskien kyselytutkimusta. Kommenttien perusteella

kysely oli hyvin strukturoitu, mutta hieman liian pitkä.

Monimuuttujamenetelmien käyttö analyyseissa osoittautui vähintäänkin haastavaksi, sillä muuttujat eivät noudattaneet jakaumiltaan menetelmien vaatimia oletuksia. Faktoriaanalyysin teossa jouduin siirtymään konfirmatorisesta lähestymistavasta eksploratiiviseen, sillä muuttujat eivät latautuneet valmistelemani mallin mukaisesti. Faktorit selittivät yhdessä vain 32 % vaihtelusta ja 68 % jäi näin mittausvirheille, mikä on astetta huonompi tulos kuin yleensä kyselytutkimuksissa (n. 50 %). Lisäksi kysymysten kommunaliteetit rotatoidussa latausmatriisissa olivat alhaisia kysymysten vaihtelun ollessa pientä.

Analyysin tuloksena saatuja vastaajakohtaisia faktoripisteitä käytettiin ryhmittelyanalyysin tekoon, jotta joukosta löydettäisiin opiskelijajoukkoja- tai tyyppejä. Oletuksena oli, että aineistosta löytyisi ainakin rikkaiden aasialaisten ryhmä, jolle raha ja maksullisuus ei merkitse mitään sekä köyhempi aasialaisten ryhmä, jotka nimenomaan ovat tulleet maksuttomuuden perässä Suomeen. Ryhmien lukumäärää tunnusteltiin kokeilemalla hierarkista ryhmittelyä osa-aineistolle käyttäen kahta erilaista ryhmittelymenetelmää. Lopulliset ryhmät muodostettiin koko aineistosta laskennallisesti tehokkaammalla ei-hierarkisella menetelmällä ryhmien lukumäärän (4) perustuessa edelliseen menetelmään. Tämän ryhmittelyn pohjalta ryhmien ominaisuuksia analysoitiin ryhmäkeskusten sekä taustamuuttujien avulla. Muodostin saatavilla olevan informaation perusteella ryhmille nimet, jotka pyrkivät kuvaamaan niitä. Ainakin rikkaampien maksuttomuuteen negatiivisesti suhtautuvien ryhmä löytyi.

Tulkinta jäi näissä monimuuttuja-analyyseissä siis pintapuoliseksi, sillä alkuperäinen vertailtavuus ei ollut mahdollista ja muuttujat soveltuivat huonosti menetelmien vaatimiin teoriakehyksiin. Menetelmien empiirisiin tuloksiin tulee siis suhtautua esimerkinomaisina.

Pilottitutkimuksena tämä tutkimus oli onnistunut ja sen avulla saatiin paljon tietoa kyselynteosta sekä opiskelijoista ja heidän taustoistaan. Tutkimusta uusittaessa on syytä käyttää lyhyempää kyselylomaketta ja ottaa paremmin huomioon kaikki erityisryhmät ulkomaalaisten tutkinto-opiskelijoiden joukossa. Heillä kaikilla on hyvin erilainen kuva opiskeluista, sen maksuista sekä opiskelun kuluista. Opiskelijat olivat tämän tutkimuksen mukaan hyvin samanmielisiä koulutuksen maksuttomuudesta, mikä voi johtua joko homogeenisestä joukosta ja siitä että Suomeen hakeutuu maksuttomuutta arvostavia opiskelijoita, tai vääristyneistä tuloksista heikon kysymyksenasettelun vuoksi. Kysymyksiä tulisi muotoilla niin, että ne käsittelisivät aihetta vielä monipuolisemmin sekä tarkemmin pohjautuen vankempaan teoriaan. Lisäksi avoimet kentät tuottivat paljon pitkiäkin vastauksia, jotka tulen antamaan AYY:lle jatkoanalyysijä varten.

Tutkimuksen tuloksia voi sinällään käyttää hyväksi opiskelijoiden edunvalvontatyössä myös muun kuin lukukausimaksukokeilun parissa. Esimerkiksi ulkomaalaisten opiskelijoiden sosioekonomista taustaa tai rahankäyttöä ei ole ennen tätä tutkittu vastaavalla laajuudella (henkilökohtainen viestintä, Hanna Sauli, syksy 2011). Lisäksi tämän pilottitutkimuksen ja sen tulosten perusteella AYY:llä on mahdollisuus toteuttaa seuraavina vuosina vastaavanlainen kysely omille opiskelijoilleen, tai laajentaa tutkimusta koko Suomea koskevaksi, ja saada näin parempaa objektiivista tietoa koskien lukukausimaksukokeilua ja kokeilun mahdollisia vaikutuksia opiskelija-ainekseen.

Lähteet

- (2007). Pääministeri Matti Vanhasen toisen hallituksen ohjelma. <http://www.valtioneuvosto.fi/tietoarkisto/aiemmat-hallitukset/vanhanenII/hallitusohjelma/pdf/hallitusohjelma-painoversio-040507.pdf>, luettu 25.8.2011.
- (2009). Yliopistolaki. <http://www.finlex.fi/fi/laki/alkup/2009/20090558>, luettu 2.1.2012.
- Aalto Working Group (2010). Aalto working group on fees and scholarships: Report. Henkilökohtainen viestintä, Hanna Sauli, toukokuu 2011.
- Aalto yliopisto (2010). Lukukausimaksukokeilun seuranta ja arviointi, kysely osa 1, aalto yliopisto, määräaika 31.12.2010. Henkilökohtainen viestintä, Hanna Sauli, toukokuu 2011.
- Aalto-yliopiston hakijapalvelut (2011). Aalto-yliopistoon yli 14 500 hakijaa kevään hauissa. <http://www.aalto.fi/fi/current/news/view/2011-04-21-003/>, luettu 2.1.2012.
- Aalto-yliopiston ylioppilaskunta, AYY (2010). Linjapaperi: Lukukausimaksukokeilun hyödyt tulee osoittaa. <http://ayy.fi/blog/2010/03/18/linjapaperi-lukukausimaksukokeilusta/>, luettu 10.7.2011.
- Ala-Panula, Riina; Bailey, Kristiina; Löppönen, Piia & Rautanen, Anne (2009). *Opiskelijarahojen käyttöönoton vaikutukset Suomessa*. Tampere University Press.
- Asplund, Rita (2006). *Koulutuksen taloudelliset vaikutukset*. SITRA.
- Asplund, Rita; Abdelkarim, Oussama Ben & Skalli, Ali (2009). Student loans and the likelihood of graduation: Evidence from finnish cohort data. *Higher Education in Europe*, 34.
- Baker, Reginald; Crawford, Scott & Swinehart, Janice (2004). *Development and Testing of Web Questionnaires*, sivut 361–384. Wiley.
- Barr, Nicholas (2004). Higher education funding. *Oxford review of Economic Policy*, 20, 264–283.
- Börsch-Supan, Axel; Elsner, Detlev; Fassbender, Heino; Kiefer, Rainer; McFadden, Daniel & Winter, Joachim (2004). How to make internet surveys representative: A case study of a two-step weighting procedure.
- Cary, NC: SAS Institute (2011). *SAS/STAT 9.3 User's Guide. Chapter 30*. SAS Institute.
- Chapman, Bruce & Ryan, Chris (2005). The access implications of income-contingent charges for higher education: lessons from Australia. *Economics of Education Review*, 24, 491–512.
- Coelli, Michael B. (2009). Tuition fees and equality of university enrolment. *Canadian Journal of Economics*, 42, 1072–1099.
- Couper, Mick P. (2008). *Designing effective Web surveys*. Cambridge University Press.
- Couper, Mick P.; Peytchev, Andy; McCabe, Sean E. & Crawford, Scott D. (2006). Web survey design - pagin versus scrolling. *Public Opinion Quarterly*, 70, 596–607.
- Curtin, Tim (2000). All taxes are graduate taxes. *The Round Table*, 356.

- de la Fuente, Angel (2003). Human capital in a global and knowledge-based economy, part II: assessment at the EU country level.
- Dillman, Don A. (1991). The design and administration of mail surveys. *Annual Review of Sociology*, 17, 225–249.
- Dillman, Don A. (2000). *Mail and Internet Surveys*. Wiley.
- Dillman, Don A. (2009). *Internet, mail and mixedmode surveys*. Wiley.
- Eicher, Jean-Claude (1998). The cost of financing of higher education in europe. *European Journal of Education*, 33, 31–39.
- Euroopan komissio (2011). Higher education in europe. http://ec.europa.eu/education/lifelong-learning-policy/doc62_en.htm, luettu 2.1.2012.
- Groves, Robert M.; Fowler, Flyoyd J. Jr.; Couper, Mick P. & Lepkowski, James M. (2004). *Survey Methodology*. Wiley.
- Harkness, Janet; Pennell, Beth-Ellen & SchouaGlusberg, Alisù (2004). *Survey Questionnaire Translation and Assessment*, sivut 453–473. Wiley.
- Henttonen, Aleks (2008). *Valinnanvapaus, vastuu ja tasa-arvo - Koulutuksen kustannusten ja kaminen*. Kalevi Sorsa -säätö.
- Häkkinen, Iida & Uusitalo, Roope (2003). The effect of a student aid reform on graduation: a duration analysis. *Uppsala University*.
- Hölttä, Seppo (1998). The funding of universities in Finland: Towards goal-oriented government steering. *European Journal of Education*, 33, 55–63.
- Johnson, Richard A. & Wichern, Dean W. (2007). *Applied Multivariate Statistical Analysis*. Pearson Education.
- Johnstone, Bruce (2004). The economics and politics of cost sharing in higher education: comparative perspectives. *Economics of Education Review*, 23, 403–410.
- Jöreskog, Karl G. (2004). Factor analysis and its extensions. <http://www.fal00.info/joreskog.pdf>, luettu 10.12.2011.
- Kaufman, L. & Rousseeuw, P. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis. 2nd edition*. Wiley.
- Kepa ry (2011). Kepa uutiset: Lukukausimaksut herättävät närää järjestökentällä, 3.2.2011. <http://www.kepa.fi/uutiset/8151>, luettu 2.1.2012.
- KvantiMOTV (2004). Yhteiskuntatieteellinen tietoarkisto. menetelmäopetuksen tietovaranto. faktorianalyysi. <http://www.fsd.uta.fi/metelmaopetus/faktori/faktori.html>, luettu 2.1.2012.
- Laaksonen, Seppo (2009). Surveymetodiikka, kurssimateriaali. Matematiikan ja tilastotieteen laitos, Helsingin yliopisto.
- Lensvelt-Mulders, G. (2008). *Surveying Sensitive Topics*, sivut 461–478. Psychology Press, Taylor & Francis Group.
- Lukukausimaksukokeilun seuranta- ja arviointityöryhmä (2010). Korkeakoulujen lukukausimaksukokeilun seuranta- ja arviointisuunnitelma 2010–2014. http://www.minedu.fi/export/sites/default/OPM/Koulutus/artikkelit/lukukausimaksukokeilu/liitteet/Kirje_10.6.2010_LIITE_Seuranta_ja_arviointisuunnitelma.pdf, luettu 5.12.2011.
- Lukukausimaksukokeilun seuranta- ja arviointityöryhmä (2011). Korkeakoulujen lukukausimaksukokeilun seuranta ja arviointi. Työryhmän väliraportti I. <http://www.minedu.fi/OPM/Koulutus/artikkelit/lukukausimaksukokeilu/index.html>, luettu 5.12.2011.

- Lynn, Peter (2008). *The problem of nonresponse*, sivut 35–55. Psychology Press, Taylor & Francis Group.
- Malhotra, Naresh K. & Birks, Davis F. (2007). *Marketing Research. An Applied Approach*. Prentice Hall.
- Marcucci, Pamela & Johnstone, Bruce (2007). Tuition fee policies in a comparative perspective: Theoretical and political rationales. *Journal of higher education Policy and Management*, 29.
- Mardia, K.V.; Kent, J.T & Bibby, J.M (1989). *Multivariate Analysis*. Academic Press.
- Mellin, Ilkka (2007). Ryhmittelyanalyysi. http://www.sal.tkk.fi/vanhat_sivut/Opinnot/Mat-2.112/K07/CLUST10.pdf, luettu 2.2.2012.
- Mellin, Ilkka (2009). Faktorianalyysi. <http://www.sal.tkk.fi/Opinnot/Mat-2.112/K07/FACTO10.pdf>, luettu 10.12.2009.
- Mustonen, Seppo (1995). *Tilastolliset monimuuttujamenetelmät*. Survo Systems.
- Mustonen, Seppo (1996). *Survo ja minä*. Survo Systems.
- Nummenmaa, L. (2004). *Käyttäytymistieteiden tilastolliset menetelmät*. Tammi.
- OECD (2007a). Education at glance 2007. www.oecd.org/edu/eag2007, luettu 30.11.2011.
- OECD (2007b). Oecd briefing note for finland 2007. <http://www.oecd.org/dataoecd/24/52/39315711.pdf>, luettu 30.11.2011.
- OECD (2009). Education at glance 2009. www.oecd.org/edu/eag2009, luettu 30.11.2011.
- OECD (2010). Education at glance 2010. www.oecd.org/edu/eag2010, luettu 30.11.2011.
- OECD (2011). Education at glance 2011. www.oecd.org/edu/eag2011, luettu 30.11.2011.
- Opetusministeriö (2005). Korkeakoulujen ulkomaisten tutkinto-opiskelijoiden maksutyöryhmän muistio. http://www.minedu.fi/OPM/Julkaisut/2005/korkeakoulujen_ulkomaisten_tutkinto-opiskelijoiden_maksutyoryhma, luettu 25.8.2011.
- Opetusministeriö (2008). Koulutus ja tutkimus 2007–2012, kehittämissuunnitelma.
- Opetusministeriö (2010). Ei paikoillanne, vaan valmiit, hep! <http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2010/liitteet/tr11.pdf?lang=fi>, luettu 10.6.2011.
- Opetusministeriö (2011). Korkeakoulujen lukukausimaksukokeilu. <http://www.minedu.fi/OPM/Koulutus/artikkelit/lukukausimaksukokeilu/index.html>, luettu 5.12.2011.
- Puustelli, Pia; Niskanen, Rebekka; Sund, Anne-Mari & Rajala, Carina (2009). *Tuleeko korkeakoulutuksesta hyödyke ja opiskelijoista asiakkaita?* Tampere University Press.
- Sauli, Hanna (2011). Ayy blogi: Lukukausimaksukokeilu, pelot ja toiveet. <http://ayy.fi/blogi/2011/06/01/lukukausimaksukokeilu-pelot-ja-toiveet/>, luettu 30.11.2011.
- Schwarz, Norbet; Knäuper, Bärbel; Oyserman, Daphna & Stich, Christine (2008). *Psychology of Asking Questions*, sivut 18–34. Psychology Press, Taylor & Francis Group.
- Teeri, Tuula (2010). Rehtorin päätös: Lukukausimaksukokeiluun otettavat maisteriohjelmat. Henkilökohtainen viestintä, Hanna Sauli, toukokuu 2011.

- Tilastokeskus (2007). Laatusuhteet tilastoissa. http://www.stat.fi/meta/qg_2ed.pdf, luettu 2.1.2012.
- Tilastokeskus (2011). Suomen virallinen tilasto (SVT): Tieto- ja viestintätekniikan käyttö [verkojulkaisu]. <http://www.stat.fi/til/sutivi/>, luettu 5.12.2011.
- Ulkoministeriön kehitysviestintä (2011). global.finland.fi: Yliopistot aloittivat varovasti. <http://www.global.finland.fi/public/default.aspx?contentid=222477>, luettu 2.3.2012.
- Unesco (2009). Trends in global higher education: Tracking an academic revolution. <http://unesdoc.unesco.org/images/0018/001831/183168e.pdf>, luettu 30.11.2011.
- Usher, A. & Cervenán, A. (2005). Global higher education rankings.
- Vehkalahti, Kimmo (2008). *Kyselytutkimuksen mittarit ja menetelmät*. Tammi.
- Vossensteyn, Hans (2009). Challenges in Student Financing: State Financial Support to Students. A worldwide perspective. *Higher Education in Europe*, 34, 171–187.
- YLE (2011). Lukukausimaksukokeilu alkaa nihkeästi, 19.5.2011. http://yle.fi/alueet/savo/2011/05/lukukausimaksukokeilu_alkaa_nihkeasti_2603020.html, luettu 2.1.2012.

A Alustava tutkimussuunnitelma

Tutkimussuunnitelma

1. Tutkimusaihe ja tutkimusongelma: mitä ja miksi?

Tutkimuksen tavoitteena on selvittää vaikuttaako Suomessa nyt käynnissä oleva lukukausimaksukokeilu EU- ja ETA-maista tulevien opiskelijoiden sosioekonomiseen taustaan. Aalto-yliopisto lähti lukukausimaksukokeiluun syksyllä 2011 ottaen käyttöön 8 000 euron lukukausimaksut 14 englanninkielisessä maisteriohjelmassa. Lukukausimaksukokeiluun kuuluu opetusministeriön asettamana pakollisena stipendijärjestelmä, jotta kaikilla hyväksytyillä olisi mahdollisuus opiskella. Aalto-yliopistossa stipendit jaettiin akateemisen suoritusten perusteella.

Tutkimus on Aalto-yliopiston ylioppilaskunnan (AYY) tilaama. Ylioppilaskuntien yleisenä huolena on opiskelija-aineiden voimakas vinoutuminen maksujen seurauksena. Toinen huolestuttava seikka on tasa-arvo maksavien ja ilmaiseksi opiskelevien opiskelijoiden välillä. Ylioppilaskunnat ovat useasti ottaneet kantaa lukukausimaksuasiaan (HYY, 2007). AYY:lle suurin motivoiva tekijä tutkimuksen tilaamiseen oli tieto siitä, ettei lukukausimaksukokeilua monitoroiva työryhmä aio selvittää sosioekonomisia vaikutuksia ollenkaan.

2. Teoreettiset lähtökohdat

Kirjallisuudessa, varsinkin kansantaloustieteellisessä (mm. Barr, 2004), on valloilla näkemys, että lukukausimaksut ovat välttämättömät korkeakoulutuksen rahoituspohjan laajentamiseen. Useat järjestöt kuten OECD ja Elinkeinoelämän keskusliitto (EK) ovat kampanjoineet aktiivisesti lukukausimaksujen puolesta, ja useat Euroopan maat ovat siirtyneet ainakin osaksi maksullisuuteen. Lisäpainetta Suomeen on tuonut muiden pohjoismaiden, Tanskan ja Ruotsin, siirtyminen osittaiseen maksullisuuteen korkeakouluissaan.

Kuitenkin pienienkin maksujen esittäminen on todistettavasti pudottanut alemman sosioekonomisen taustan omaavien osuutta korkeakoulutukseen osallistuvista (Coelli, 2004). Korkeampaan sosioekonomiseen luokkaan maksujen esittäminen ei ole vaikuttanut merkittävästi.

3. Metodi

Sosioekonomisen taustan ja muita muutoksia tutkitaan opiskelijoille lähetettävän kyselylomakkeen avulla. Tarkoituksena on verrata 2010 ja 2011 samoissa ohjelmissa aloittaneita EU- ja ETA-maiden ulkopuolisia opiskelijoita toisiinsa.

Tutkimuksen edetessä kävi kuitenkin niin, että maksuohjelmiin hyväksytyistä opiskelijoista moni perui opiskelupaikkansa. Opiskelijoiden eli havaintojen vähyys aiheutti ongelmia, joten tutkimusta piti laajentaa koskemaan kaikkia opiskelijoita englanninkielisissä maisteriohjelmissa.

Tuloksia on tarkoitus analysoida monimuuttujamenetelmillä datan keruun sekä tarkastelun jälkeen.

4. Aineisto

Aineisto kerätään kyselylomakkeella. Kyselylomakkeessa mitataan motivaatiota, asennetta, sosioekonomista taustaa sekä muita tausta tekijöitä. Seuraavassa esimerkkejä jokaisesta kysymyslajista:

Motivaatiota mitataan väittämällä, kuten " High tuition fees guarantee a high quality of education". Väittämiä saa arvioida viisiportaisella asteikolla "I totally disagree" --- "I totally agree". Motivaatiota opiskella Suomessa/Aalto-yliopistossa mitataan esittämällä monia kysymyksiä ja pyytäen arvioimaan niitä ei yhtään tärkeästä tärkeään. Yksi esimerkki väittämä on " Good reputation of the university". Sosioekonomista taustaa mitataan mm. kysymyksellä " Some people are considered to have a high social standing and some are considered to have a low social standing. Thinking about your family background, where would you place your parents on this scale?". Taustatiedoissa kysytään mm. kotimaata sekä ikää ja sukupuolta.

5. Tutkimuksen tulokset ja merkitys

Datan vähyys asettaa isoja esteitä tulosten kunnolliseen tulkintaan.

Kyselylomaketta voi kuitenkin käyttää tehokkaasti myös ensi vuonna kun useammat korkeakoulut lähtevät maksukokeiluun mukaan.

Lisäksi Opiskelijajärjestöjen tutkimussäätiö (OTUS) voisi käyttää lomaketta ja/tai tutkimuksen tuloksia omassa työssään.

6. Aikataulu

Aloitin työn tekemisen kesäkuussa 2011 hakemalla ensin tietoa kirjallisuudesta sekä ryhtyen muotoilemaan kyselylomaketta.

Kappaleet ilmiön taustasta sekä kyselylomakkeen teosta muotoutuivat elo-lokakuussa 2011, ja niitä on tarkoitus jatkaa datankeräysvaiheessa.

Datan keräys oli suunniteltu tapahtuvan lokakuun 2011 aikana, mutta se myöhästyi tietosuojasyistä. Näin ollen myös tulosten analysointi voidaan aloittaa aikaisintaan joulukuun alussa ja työ on valmis todennäköisesti tammi-helmikuussa 2011.

Lähteet:

Barr, Nicholas (2004). *Oxford Review of Economic Policy*. Higher education funding.

Coelli, Michael (2004). Tuition Increases and Inequality in Post-Secondary Education Attendance. Canadian Economics Association Meetings in Toronto

Helsingin yliopiston ylioppilaskunta (HYY). 6 väitettä ja faktaa lukukausimaksuista.

<https://hyy.helsinki.fi/tiedostot/6%20vaitetta.pdf> luettu 27.10.2011

B Sitoumus

SITOUMUS

1 (1)

HENKILÖTIETOJEN LUOVUTUKSENSAAJAN SITOUMUS KOSKIEN JULKISUUSLAIN 16.3 §:N PERUSTEELLA LUOVUTETTETTUIEN TIETOJEN KÄSITTELYÄ

Luovuttaja	Aalto-yliopisto
Luovutuksensaaja	Eija-Leena Koponen
Rekisteri	Aalto-yliopiston opiskelijatietojärjestelmä
Luovutettavat tiedot	Lukuvuosimaksullisiin maisteriohjelmiin 2011 valittujen opiskelijoiden nimet ja sähköpostiosoitteet
Käsittelyn peruste	Tieteellinen tutkimus (henkilötietolaki 14 §)

Allekirjoittanut luovutuksensaaja sitoutuu käyttämään saamiaan henkilötietoja ainoastaan 5.10.2011 päivätyn luovutuspyyntönsä mukaiseen käyttötarkoitukseen.

Luovutuksensaaja sitoutuu käsittelemään saamiaan henkilötietoja 5.10.2011 päivätyn luovutuspyyntönsä mukaisesti vain henkilökohtaisesti ja olemaan luovuttamatta niitä eteenpäin.

Luovutuksensaaja sitoutuu käsittelemään saamiaan henkilötietoja huolellisesti ja suojaamaan ne hyvin siten, että ne pysyvät virheettöminä ja eheinä, eivätkä voi joutua ulkopuolisten saataville. Luovutuksensaaja sitoutuu hävittämään tiedot heti, kun niitä ei enää tarvita yllä mainittuun käyttötarkoitukseensa.

Paikka ja päiväys

Allekirjoitus

Eija-Leena Koponen

C Saatekirje

To: *respondent name*

Vaihtoehto 1: Subject: Free education? Survey for international students

Vaihtoehto 2: Subject: Reminder: Take the survey on tuition fees and win a 100 euro voucher!

Vaihtoehto 3: Subject: Survey for international students. We need You!

Dear international degree student,

Please fill in this questionnaire about international degree students at Aalto University. It takes only 10-15 minutes of your time and after filling the questionnaire there's a chance to take part in a lottery where you can win gift voucher of 100 euros to Stockmann!

The results will be a great help for AYY (Aalto University Student Union) in promoting the interests of students regarding the tuition fee trial. This may benefit you or all those studying in Finnish universities in the years to come. I will also submit the study as my Master's thesis to the University of Helsinki.

Please follow the link to access the survey:

#url#

Username: # tunnus#

Password: # salasana#

(In order to avoid double and false response, you need to fill in the above username and password when starting the survey.)

Thank you!

Eija-Leena Koponen

eija-leena.koponen@helsinki.fi

D Kyselylomake

Welcome to the survey for international degree students in Aalto University. This survey contains 8 pages of which others are extremely short and others a bit longer. After filling out questions, you'll end up in submitting page from where you'll be directed to the lottery page. It will take approximately 10-15 minutes to finish the questionnaire.

Survey for international degree students in Aalto University

Registry description

1 / 8

As a student, you probably have some opinions regarding tuition fees and their effects on quality, student equality and so on.
Below there are some statements I would like you to take a stance on.

	I totally disagree	I totally agree	I do not know
Free education is a basic right	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
High tuition fees guarantee a high quality of education	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Paying and non-paying students are treated equally in Aalto	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Scholarships should be given to poor students	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Scholarships should be given to the students with best academic performance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The differences in quality between the free and chargeable master degree programs in Aalto are high	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tuition fees prevent poor students from pursuing higher education	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tuition fees should be used to improve the quality of teaching, even if non-paying students also benefit from it	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Any other opinions regarding tuition fees?

Next >>

Survey for international degree students in Aalto University

[Registry description](#)

2 / 8

In the next questions you should choose how important given factors were in your choice of study destination. If you find these alternatives too strict or not applicable, please feel free to write in the empty space below.

How important were the following factors in your decision to move in Finland?

	Not important	Very important	I do not know
Career development/work opportunities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Cost of study	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Experiencing a foreign country	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Family ties (incl. boy/girlfriend)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hardship in home country	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
High standard of living in Finland	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Learning Finnish/Swedish	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Possible migration/citizenship	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Prior knowledge of Finland	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pursuing a university degree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reputation of Finnish education	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reputation of Finnish industry	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Visa Processing system	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If you chose other, please tell about it

How important were the following factors in your decision to study at the Aalto University?

	Not important	Very important	I do not know
Academic freedom/independence	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Employment chances after studies	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Good reputation of the university	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Good reputation of my study program	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Possibility of studying in English	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality of research	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality of student facilities	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality of teaching	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Recommendation (from a friend for example)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reputation (value in my career) of a qualification from this university	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Scholarship	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Social life	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
University's ranking position	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

If you chose other, please tell about it

What are you going to do after your studies at Aalto University?
Please select from below. ?

Survey for international degree students in Aalto University

[Registry description](#)

3 / 8

Now I would like you to fill in some background information regarding your education and the enrollment to Aalto University.

This information is used to do comparative analysis. Thus the data from this question as well as from the other questions, will be used in such a way that individuals cannot be recognized.

Nationality ?

--Select--

Please select your enrolment year to Aalto University i.e. the year you started your current studies. ?

- ☐ 2010
☐ 2011
☐ Other

Please select your master's degree program at Aalto University. Please note that you first (1) have to choose the school, then (2) the program.

--Select--

If other, please specify

Were you accepted in any other university than Aalto?

- ☐ No
☐ Yes
☐ I do not wish to tell

<< Previous

Next >>

Survey for international degree students in Aalto University

[Registry description](#)

4 / 8

Aalto University is taking part in a tuition fee experiment in Finland. In this experiment some of the English Master's degrees charge tuition fees (8 000 euros/year) from non EU/ETA students.

Please select from below whether you are being charged fees or not. These selections are used only to compare results between groups.

Do you pay tuition fees / get scholarship from Aalto?

- ☐ Yes
- ☐ No
- ☐ I do not know

If yes, please select one of the following category

- I get from Aalto
 - ☐ a full scholarship covering the tuition fee and living costs
 - ☐ a full scholarship covering the tuition fee
 - ☐ 50% of a full scholarship covering the tuition fee
 - ☐ no scholarship, I pay the full fee
 - ☐ Not applicable

<< Previous

Next >>

Page 4 / 10

© Eduix Oy

Survey for international degree students in Aalto University

[Registry description](#)

5 / 8

Below there are some statements I would like you to take a stance on.

	I totally disagree	I totally agree	I do not know
My master's degree is big investment for my family when it comes to money	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My master's degree is big investment for myself when it comes to money	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My family took a loan to cover my study costs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I took a personal loan to cover my study costs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My family sold property to cover my study costs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I sold some of my property to cover my study costs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My family works extra hard to cover my expenses while studying	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I had/have to work extra hard to cover my expenses while studying	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Feel free to comment

~~Some people are considered to have a high social standing and some are considered to have a low social standing. Thinking about your family~~

- ☐ High social standing
- ☐
- ☐
- ☐
- ☐
- ☐
- ☐
- ☐
- ☐
- ☐ Low social standing
- ☐ Do not know

Parents' education

- | | |
|--|--|
| The highest degree that my mother possesses is | <input type="radio"/> No degree |
| | <input type="radio"/> Up to lower secondary |
| | <input type="radio"/> Upper secondary (e.g. High school or equivalent) |
| | <input type="radio"/> Bachelor or equivalent |
| | <input type="radio"/> Master or equivalent |
| | <input type="radio"/> Doctoral (PhD) |
| | <input type="radio"/> Other |
| | <input type="radio"/> Do not know/ Not applicable |
| The highest degree that my father possesses is | <input type="radio"/> No degree |
| | <input type="radio"/> Up to lower secondary |
| | <input type="radio"/> Upper secondary (e.g. High school or equivalent) |
| | <input type="radio"/> Bachelor or equivalent |
| | <input type="radio"/> Master or equivalent |
| | <input type="radio"/> Doctoral (PhD) |
| | <input type="radio"/> Other |
| | <input type="radio"/> Do not know/ Not applicable |

Survey for international degree students in Aalto University

[Registry description](#)

6a / 8

Estimation of your monthly expenses in Finland in euros (incl. Rent, food, transportation etc.) ?

Please select one to two (1-2) most important sources of funding for your studies/living expenses at the moment.

My primary (1st) source of funding is

- ☐ Contribution from parents/family
- ☐ Own income from previous job
- ☐ Part-time working
- ☐ Study grants/loans from Finland
- ☐ Support by home state loan (repayable)
- ☐ Support by home state grant (non-repayable)
- ☐ Charity or other trust
- ☐ Other loan
- ☐ Other

(Optional) My secondary (2nd) source of funding is

- ☐ Contribution from parents/family
- ☐ Own income from previous job
- ☐ Part-time working
- ☐ Study grants/loans from Finland
- ☐ Support by home state loan (repayable)
- ☐ Support by home state grant (non-repayable)
- ☐ Charity or other trust
- ☐ Other loan
- ☐ Other

If you chose other, please feel free to tell about it.

I would have chosen Aalto University as a place for studies even if I had to pay tuition fees.

I totally disagree

☐

.....

☐

.....

☐

.....

☐

I totally agree

☐

Why? Please tell about it.

If yes how much would you be willing to pay for your studies? (in euros per academic year)

<< Previous Next >>

Survey for international degree students in Aalto University

[Registry description](#)

6b / 8

Estimation of your monthly expenses in Finland in euros (incl. Rent, food, transportation etc. BUT excluding tuition fees) ?

Please select one to two (1-2) most important sources of funding for your studies/living expenses at the moment.

My primary (1st) source of funding is

- ☐ Contribution from parents/family
- ☐ Own income from previous job
- ☐ Scholarship from Aalto University
- ☐ Part-time working
- ☐ Study grants/loans from Finland
- ☐ Support by home state loan (repayable)
- ☐ Support by home state grant (non-repayable)
- ☐ Charity or other trust
- ☐ Other loan
- ☐ Other

(Optional) My secondary (2nd) source of funding is

- ☐ Contribution from parents/family
- ☐ Own income from previous job
- ☐ Scholarship from Aalto University
- ☐ Part-time working
- ☐ Study grants/loans from Finland
- ☐ Support by home state loan (repayable)
- ☐ Support by home state grant (non-repayable)
- ☐ Charity or other trust
- ☐ Other loan
- ☐ Other

If you chose other, please feel free to tell about it.

If Aalto University did not charge tuition fees, I would not have chosen to study at Aalto.

I totally
disagree

☐

.....

☐

.....

☐

.....

☐

I totally
agree

☐

I do not know

☐

Why? Please tell about it.

[<< Previous](#) [Next >>](#)

Survey for international degree students in Aalto University

[Registry description](#)

7 / 8

Finally, please fill in some personal information. This information is used to do comparative analysis and get the idea of the overall age, gender etc. status of respondents. In other words data will be used in such a way that individuals cannot be recognized.

After this you have successfully completed the survey!

Year of birth (YYYY). ?

Gender ?

- ☐ Female
☐ Male
☐ I do not wish to tell

The year you moved to Finland. ?

[<< Previous](#) [Next >>](#)

Survey for international degree students in Aalto University

[Registry description](#)

8 / 8

I greatly thank You for participating in this survey! Without your answer my research wouldn't be complete.

If you have anything you wish to comment or tell, please feel free to use this space. The stage is yours!

[<< Previous](#) [Next >>](#)

In the next page, after submitting the form, you'll be directed to the page where you can take part in the lottery!!

E Perustunnuslukuja

Tämä on muokkaamaton data suoraan lomakkeesta.
99= en tiedä

Variable: V_1 "Free education is a basic right"
N(missing)=2
min=1 in obs.#40
max=99 in obs.#16
mean=6.549618 stddev=14.20686 skewness=6.320497 kurtosis=38.17343
lower_Q=4 median=5 upper_Q=5
V_1 f % *=8 obs.
1 12 3.1 *
2 13 3.3 *
3 40 10.2 *****
4 70 17.8 *****
5 249 63.4 *****
99 9 2.3 *

Variable: V_2 "High tuition fees guarantee a high quality of education"
N(missing)=4
min=1 in obs.#2
max=99 in obs.#78
mean=2.831202 stddev=9.841938 skewness=9.565943 kurtosis=90.50024
lower_Q=1 median=2 upper_Q=2
V_2 f % *=4 obs.
1 184 47.1 *****
2 123 31.5 *****
3 51 13.0 *****
4 17 4.3 ****
5 12 3.1 ***
99 4 1.0 *

Variable: V_3 "Paying and non-paying students are treated equally in Aalto"
N(missing)=5
min=1 in obs.#25
max=99 in obs.#2
mean=39.58462 stddev=45.77464 skewness=0.5289 kurtosis=-1.721446
lower_Q=5 median=5 upper_Q=99
V_3 f % *=4 obs.
1 8 2.1 **
2 11 2.8 **
3 16 4.1 ****
4 45 11.5 *****
5 165 42.3 *****
99 145 37.2 *****

Variable: V_4 "Scholarships should be given to poor students "

N(missing)=3

min=1 in obs.#10

max=99 in obs.#2

mean=7.170918 stddev=17.73934 skewness=4.9569 kurtosis=22.71385

lower_Q=3 median=4 upper_Q=5

V_4	f	%	*=4 obs.
1	27	6.9	*****
2	41	10.5	*****
3	75	19.1	*****
4	84	21.4	*****
5	151	38.5	*****
99	14	3.6	***

Variable: V_5 "Scholarships should be given to the students with best academi

N(missing)=5

min=1 in obs.#16

max=99 in obs.#2

mean=5.348718 stddev=10.73573 skewness=8.526683 kurtosis=71.43175

lower_Q=4 median=4 upper_Q=5

V_5	f	%	*=4 obs.
1	11	2.8	**
2	22	5.6	*****
3	52	13.3	*****
4	120	30.8	*****
5	180	46.2	*****
99	5	1.3	*

Variable: V_6 "The differences in quality between the free and chargeable mas

N(missing)=5

min=1 in obs.#10

max=99 in obs.#1

mean=55.99744 stddev=48.23121 skewness=-0.227755 kurtosis=-1.949436

lower_Q=2 median=99 upper_Q=99

V_6	f	%	*=8 obs.
1	82	21.0	*****
2	43	11.0	*****
3	18	4.6	**
4	16	4.1	**
5	14	3.6	*
99	217	55.6	*****

Variable: V_7 "Tuition fees prevent poor students from pursuing higher educat

N(missing)=3

min=1 in obs.#23

max=99 in obs.#31

mean=5.971939 stddev=12.60563 skewness=7.189319 kurtosis=50.10482

lower_Q=4 median=5 upper_Q=5

V_7	f	%	*=8 obs.
1	16	4.1	**
2	23	5.9	**
3	21	5.4	**
4	102	26.0	*****
5	223	56.9	*****
99	7	1.8	:

```

Variable: V_8          "Tuition fees should be used to improve the quality of
teaching
N(missing)=7
min=1                in obs.#21
max=99               in obs.#15
mean=8.71134         stddev=20.5518  skewness=4.14908  kurtosis=15.29233
lower_Q=4            median=5         upper_Q=5
V_8                  f          %          *=8 obs.
    1          29      7.5 ***
    2          26      6.7 ***
    3          35      9.0 ****
    4          82     21.1 *****
    5         197     50.8 *****
    99         19      4.9 **

```

```

*****
*****

```

```

Variable: FI_1        "Career development/work opportunities"
N(missing)=4
min=1                in obs.#7
max=99               in obs.#2
mean=5.483376        stddev=11.76017  skewness=7.728011  kurtosis=58.51846
lower_Q=3            median=5         upper_Q=5
FI_1                  f          %          *=8 obs.
    1          33      8.4 ****
    2          27      6.9 ***
    3          40     10.2 *****
    4          82     21.0 *****
    5         203     51.9 *****
    99          6      1.5 :

```

```

Variable: FI_2        "Cost of study"
N(missing)=6
min=1                in obs.#16
max=99               in obs.#124
mean=5.210797        stddev=9.641608  skewness=9.481346  kurtosis=89.31151
lower_Q=4            median=5         upper_Q=5
FI_2                  f          %          *=8 obs.
    1          24      6.2 ***
    2          12      3.1 *
    3          44     11.3 *****
    4          74     19.0 *****
    5         231     59.4 *****
    99          4      1.0 :

```

```

Variable: FI_3        "Experiencing a foreign country"
N(missing)=7
min=1                in obs.#23
max=99               in obs.#230
mean=4.621134        stddev=6.898172  skewness=13.22296  kurtosis=178.0906
lower_Q=4            median=5         upper_Q=5
FI_3                  f          %          *=8 obs.
    1          23      5.9 **
    2          17      4.4 **
    3          43     11.1 *****
    4         106     27.3 *****
    5         197     50.8 *****
    99          2      0.5 :

```

Variable: FI_4 "Family ties (incl. boy/girlfriend)"
N(missing)=6
min=1 in obs.#1
max=99 in obs.#13
mean=4.287918 stddev=13.83349 skewness=6.609755 kurtosis=42.3075
lower_Q=1 median=1 upper_Q=4
FI_4 f % *=8 obs.
1 205 52.7 *****
2 32 8.2 ****
3 39 10.0 ****
4 35 9.0 ****
5 70 18.0 *****
99 8 2.1 *

Variable: FI_5 "Hardship in home country"
N(missing)=9
min=1 in obs.#4
max=99 in obs.#5
mean=5.803109 stddev=18.15625 skewness=4.907999 kurtosis=22.24709
lower_Q=1 median=2 upper_Q=4
FI_5 f % *=4 obs.
1 167 43.3 *****
2 52 13.5 *****
3 69 17.9 *****
4 44 11.4 *****
5 40 10.4 *****
99 14 3.6 ***

Variable: FI_6 "High standard of living in Finland"
N(missing)=5
min=1 in obs.#7
max=99 in obs.#2
mean=4.797436 stddev=11.87023 skewness=7.701768 kurtosis=58.1831
lower_Q=2 median=4 upper_Q=5
FI_6 f % *=4 obs.
1 62 15.9 *****
2 44 11.3 *****
3 80 20.5 *****
4 103 26.4 *****
5 95 24.4 *****
99 6 1.5 *

Variable: FI_7 "Learning Finnish/Swedish"
N(missing)=9
min=1 in obs.#1
max=99 in obs.#376
mean=2.432642 stddev=5.095643 skewness=17.68086 kurtosis=332.0235
lower_Q=1 median=2 upper_Q=3
FI_7 f % *=4 obs.
1 167 43.3 *****
2 83 21.5 *****
3 59 15.3 *****
4 50 13.0 *****
5 26 6.7 *****
99 1 0.3 :

Variable: FI_8 "Possible migration/citizenship"
N(missing)=9
min=1 in obs.#2
max=99 in obs.#38
mean=3.702073 stddev=11.0278 skewness=8.37471 kurtosis=69.42648
lower_Q=1 median=2 upper_Q=4
FI_8 f % *=4 obs.
1 154 39.9 *****
2 56 14.5 *****
3 68 17.6 *****
4 51 13.2 *****
5 52 13.5 *****
99 5 1.3 *

Variable: FI_9 "Prior knowledge of Finland"
N(missing)=5
min=1 in obs.#2
max=99 in obs.#38
mean=4.284615 stddev=11.93614 skewness=7.699984 kurtosis=58.15615
lower_Q=1 median=3 upper_Q=4
FI_9 f % *=4 obs.
1 100 25.6 *****
2 64 16.4 *****
3 89 22.8 *****
4 73 18.7 *****
5 58 14.9 *****
99 6 1.5 *

Variable: FI_10 "Pursuing a university degree"
N(missing)=5
min=1 in obs.#15
max=5 in obs.#1
mean=4.223077 stddev=1.171197 skewness=-1.60083 kurtosis=1.625309
lower_Q=4 median=5 upper_Q=5
FI_10 f % *=8 obs.
1 27 6.9 ***
2 13 3.3 *
3 32 8.2 ****
4 92 23.6 *****
5 226 57.9 *****

Variable: FI_11 "Reputation of Finnish education"
N(missing)=6
min=1 in obs.#16
max=99 in obs.#13
mean=5.041131 stddev=9.660385 skewness=9.477702 kurtosis=89.26023
lower_Q=4 median=4 upper_Q=5
FI_11 f % *=4 obs.
1 26 6.7 *****
2 17 4.4 ****
3 46 11.8 *****
4 113 29.0 *****
5 183 47.0 *****
99 4 1.0 *

Variable: FI_12 "Reputation of Finnish industry"
N(missing)=7
min=1 in obs.#4
max=99 in obs.#13
mean=5.592784 stddev=13.63076 skewness=6.644742 kurtosis=42.58694
lower_Q=3 median=4 upper_Q=5
FI_12 f % *=4 obs.
1 41 10.6 *****
2 34 8.8 *****
3 67 17.3 *****
4 122 31.4 *****
5 116 29.9 *****
99 8 2.1 **

Variable: FI_13 "Visa Processing system"
N(missing)=7
min=1 in obs.#1
max=99 in obs.#17
mean=6.020619 stddev=18.72236 skewness=4.736689 kurtosis=20.5795
lower_Q=1 median=2 upper_Q=4
FI_13 f % *=4 obs.
1 170 43.8 *****
2 56 14.4 *****
3 64 16.5 *****
4 38 9.8 *****
5 45 11.6 *****
99 15 3.9 ***

Variable: FI_14 "Other"
N(missing)=133
min=1 in obs.#13
max=99 in obs.#1
mean=64.06107 stddev=46.42404 skewness=-0.572055 kurtosis=-1.674484
lower_Q=4 median=99 upper_Q=99
FI_14 f % *=4 obs.
1 44 16.8 *****
2 3 1.1 :
3 17 6.5 ****
4 5 1.9 *
5 26 9.9 *****
99 167 63.7 *****

Variable: A_1 "Academic freedom/independence"
N(missing)=8
min=1 in obs.#12
max=99 in obs.#94
mean=8.839793 stddev=21.11454 skewness=4.019184 kurtosis=14.22994
lower_Q=3 median=5 upper_Q=5
A_1 f % *=4 obs.
1 38 9.8 *****
2 21 5.4 *****
3 49 12.7 *****
4 81 20.9 *****
5 178 46.0 *****
99 20 5.2 *****

Variable: A_2 "Employment chances after studies"
N(missing)=4
min=1 in obs.#2
max=99 in obs.#4
mean=6.795396 stddev=15.74929 skewness=5.651867 kurtosis=30.13779
lower_Q=4 median=5 upper_Q=5
A_2 f % *=8 obs.
1 26 6.6 ***
2 8 2.0 *
3 54 13.8 *****
4 96 24.6 *****
5 196 50.1 *****
99 11 2.8 *

Variable: A_3 "Good reputation of the university"
N(missing)=3
min=1 in obs.#7
max=99 in obs.#4
mean=5.961735 stddev=11.64703 skewness=7.817045 kurtosis=59.48424
lower_Q=4 median=5 upper_Q=5
A_3 f % *=8 obs.
1 9 2.3 *
2 6 1.5 :
3 28 7.1 ***
4 77 19.6 *****
5 266 67.9 *****
99 6 1.5 :

Variable: A_4 "Good reputation of my study program"
N(missing)=5
min=1 in obs.#7
max=99 in obs.#4
mean=6.4 stddev=14.2884 skewness=6.291072 kurtosis=37.81515
lower_Q=4 median=5 upper_Q=5
A_4 f % *=8 obs.
1 17 4.4 **
2 9 2.3 *
3 50 12.8 *****
4 105 26.9 *****
5 200 51.3 *****
99 9 2.3 *

Variable: A_5 "Possibility of studying in English"
N(missing)=4
min=1 in obs.#9
max=99 in obs.#110
mean=5.585678 stddev=9.55133 skewness=9.589004 kurtosis=90.84365
A_5 f % *=8 obs.
1 13 3.3 *
2 5 1.3 :
3 18 4.6 **
4 44 11.3 *****
5 307 78.5 *****
99 4 1.0 :

Variable: A_6 "Quality of research"
N(missing)=6
min=1 in obs.#23
max=99 in obs.#4
mean=9.845758 stddev=22.91911 skewness=3.623153 kurtosis=11.16821
lower_Q=3 median=4 upper_Q=5
A_6 f % *=4 obs.
1 23 5.9 *****
2 20 5.1 *****
3 57 14.7 *****
4 105 27.0 *****
5 160 41.1 *****
99 24 6.2 *****

Variable: A_7 "Quality of student facilities"
N(missing)=6
min=1 in obs.#23
max=99 in obs.#4
mean=7.688946 stddev=18.34333 skewness=4.758549 kurtosis=20.74045
lower_Q=4 median=4 upper_Q=5
A_7 f % *=4 obs.
1 20 5.1 *****
2 20 5.1 *****
3 51 13.1 *****
4 122 31.4 *****
5 161 41.4 *****
99 15 3.9 ***

Variable: A_8 "Quality of teaching"
N(missing)=6
min=1 in obs.#65
max=99 in obs.#4
mean=7.508997 stddev=17.73075 skewness=4.948216 kurtosis=22.58241
lower_Q=4 median=4 upper_Q=5
A_8 f % *=4 obs.
1 17 4.4 ****
2 13 3.3 ***
3 51 13.1 *****
4 131 33.7 *****
5 163 41.9 *****
99 14 3.6 ***

Variable: A_9 "Recommendation (from a friend for example)"
N(missing)=6
min=1 in obs.#2
max=99 in obs.#17
mean=8.437018 stddev=21.71083 skewness=3.912238 kurtosis=13.39116
lower_Q=2 median=4 upper_Q=5
A_9 f % *=4 obs.
1 78 20.1 *****
2 37 9.5 *****
3 66 17.0 *****
4 82 21.1 *****
5 105 27.0 *****
99 21 5.4 *****

Variable: A_10 "Reputation (value in my career) of a qualification"
N(missing)=8
min=1 in obs.#7
max=99 in obs.#72
mean=6.395349 stddev=14.34805 skewness=6.261881 kurtosis=37.46083
lower_Q=4 median=5 upper_Q=5
A_10 f % *=8 obs.
1 19 4.9 **
2 14 3.6 *
3 38 9.8 ****
4 112 28.9 *****
5 195 50.4 *****
99 9 2.3 *

Variable: A_11 "Scholarship"
N(missing)=10
min=1 in obs.#1
max=99 in obs.#5
mean=11.92987 stddev=28.47189 skewness=2.721573 kurtosis=5.439205
lower_Q=1 median=3 upper_Q=5
A_11 f % *=4 obs.
1 148 38.4 *****
2 27 7.0 *****
3 49 12.7 *****
4 39 10.1 *****
5 85 22.1 *****
99 37 9.6 *****

Variable: A_12 "Social life"
N(missing)=8
min=1 in obs.#2
max=99 in obs.#17
mean=9.046512 stddev=23.70879 skewness=3.517909 kurtosis=10.42686
lower_Q=2 median=3 upper_Q=4
A_12 f % *=4 obs.
1 97 25.1 *****
2 48 12.4 *****
3 87 22.5 *****
4 78 20.2 *****
5 52 13.4 *****
99 25 6.5 *****

Variable: A_13 "University's ranking position"
N(missing)=5
min=1 in obs.#3
max=99 in obs.#5
mean=7.402564 stddev=19.02128 skewness=4.581078 kurtosis=19.12128
lower_Q=3 median=4 upper_Q=5
A_13 f % *=4 obs.
1 59 15.1 *****
2 38 9.7 *****
3 58 14.9 *****
4 101 25.9 *****
5 118 30.3 *****
99 16 4.1 ****

```

Variable: A_14      Other
N(missing)=192
min=1      in obs.#13
max=99     in obs.#2
mean=63.3202  stddev=46.73764  skewness=-0.54098  kurtosis=-1.710383
lower_Q=3      median=99      upper_Q=99
A_14      f      %      *=4 obs.
    1      39      19.2 *****
    2      1       0.5 :
    3      14      6.9 ***
    4       6      3.0 *
    5      15      7.4 ***
    99     128     63.1 *****

```

```

*****
*****
Variable: VM_1      "My master's degree is big investment for my family"
N(missing)=6
min=1      in obs.#4
max=99     in obs.#6
mean=7.305913  stddev=19.07477  skewness=4.568249  kurtosis=19.0235
lower_Q=2      median=4      upper_Q=5
VM_1      f      %      *=4 obs.
    1      77      19.8 *****
    2      35      9.0 *****
    3      62      15.9 *****
    4      70      18.0 *****
    5     129      33.2 *****
    99      16      4.1 ****

```

```

Variable: VM_2      "My master's degree is big investment for myself wh"
N(missing)=9
min=1      in obs.#23
max=99     in obs.#3
mean=7.544041  stddev=18.46117  skewness=4.725633  kurtosis=20.46928
lower_Q=3      median=4      upper_Q=5
VM_2      f      %      *=4 obs.
    1      40      10.4 *****
    2      27      7.0 *****
    3      51      13.2 *****
    4      85      22.0 *****
    5     168      43.5 *****
    99      15      3.9 ***

```

```

Variable: VM_3      "My family took a loan to cover my study costs"
N(missing)=16
min=1      in obs.#1
max=99     in obs.#13
mean=4.873351  stddev=17.09651  skewness=5.289626  kurtosis=26.17003
lower_Q=1      median=1      upper_Q=2
VM_3      f      %      *=8 obs.
    1     251      66.2 *****
    2      39      10.3 *****
    3      19       5.0 **
    4      17       4.5 **
    5      41      10.8 *****
    99      12       3.2 *

```

Variable: VM_4 "I took a personal loan to cover my study costs"
N(missing)=13
min=1 in obs.#1
max=99 in obs.#13
mean=5.041885 stddev=17.01037 skewness=5.298845 kurtosis=26.3204
lower_Q=1 median=1 upper_Q=3
VM_4 f % *=8 obs.
1 239 62.6 *****
2 32 8.4 ****
3 20 5.2 **
4 20 5.2 **
5 59 15.4 *****
99 12 3.1 *

Variable: VM_5 "My family sold property to cover my study costs"
N(missing)=10
min=1 in obs.#1
max=99 in obs.#13
mean=6.049351 stddev=20.64129 skewness=4.268615 kurtosis=16.27828
VM_5 f % *=8 obs.
1 294 76.4 *****
2 22 5.7 **
3 18 4.7 **
4 10 2.6 *
5 23 6.0 **
99 18 4.7 **

Variable: VM_6 "I sold some of my property to cover my study costs"
N(missing)=15
min=1 in obs.#1
max=99 in obs.#13
mean=5.473684 stddev=19.0251 skewness=4.691892 kurtosis=20.11655
VM_6 f % *=8 obs.
1 278 73.2 *****
2 21 5.5 **
3 17 4.5 **
4 21 5.5 **
5 28 7.4 ***
99 15 3.9 *

Variable: VM_7 "My family works extra hard to cover my expenses wh"
N(missing)=10
min=1 in obs.#1
max=99 in obs.#13
mean=6.368831 stddev=19.37901 skewness=4.5382 kurtosis=18.75489
lower_Q=1 median=2 upper_Q=4
VM_7 f % *=4 obs.
1 192 49.9 *****
2 35 9.1 *****
3 32 8.3 *****
4 40 10.4 *****
5 70 18.2 *****
99 16 4.2 ****

Variable: VM_8 "I had/have to work extra hard to cover my expenses"

N(missing)=10

min=1 in obs.#6

max=99 in obs.#67

mean=6.690909 stddev=18.02604 skewness=4.88417 kurtosis=22.07444

lower_Q=1 median=4 upper_Q=5

VM_8	f	%	*=4 obs.
1	99	25.7	*****
2	36	9.4	*****
3	52	13.5	*****
4	57	14.8	*****
5	127	33.0	*****
99	14	3.6	***

Variable: Social_s Social_standing

N(missing)=3

min=0 in obs.#12

max=10 in obs.#2

mean=5.533163 stddev=2.470061 skewness=-0.584489 kurtosis=-0.062032

lower_Q=4 median=6 upper_Q=7

Social_s	f	%	*=2 obs.
0	25	6.4	*****
1	15	3.8	*****
2	6	1.5	***
3	26	6.6	*****
4	39	9.9	*****
5	53	13.5	*****
6	81	20.7	*****
7	65	16.6	*****
8	53	13.5	*****
9	12	3.1	*****
10	17	4.3	*****

Variable: Gender Gender

N(missing)=2

entropy=1.093223 (69.0%)

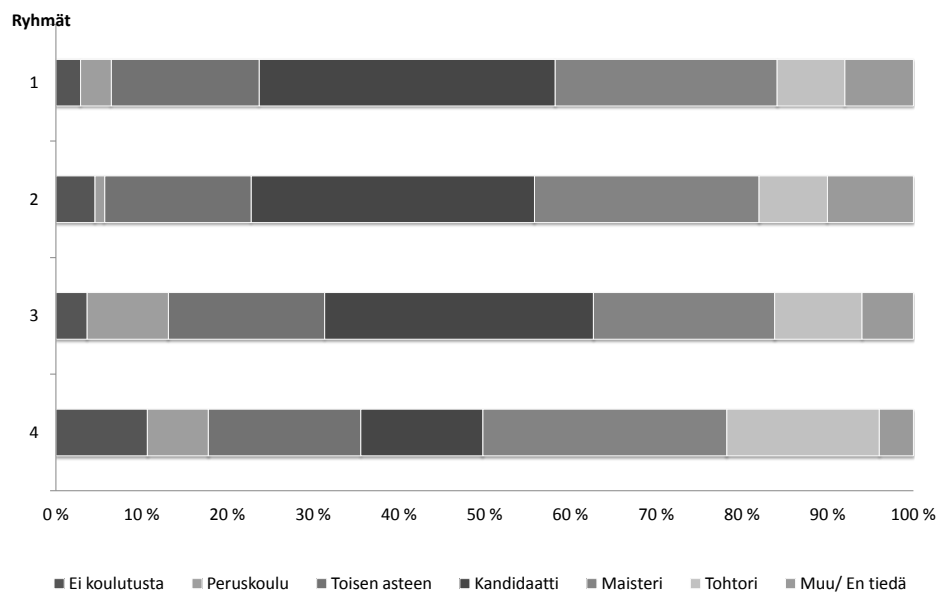
Gender f % *=8 obs.

Female 125 31.8 *****

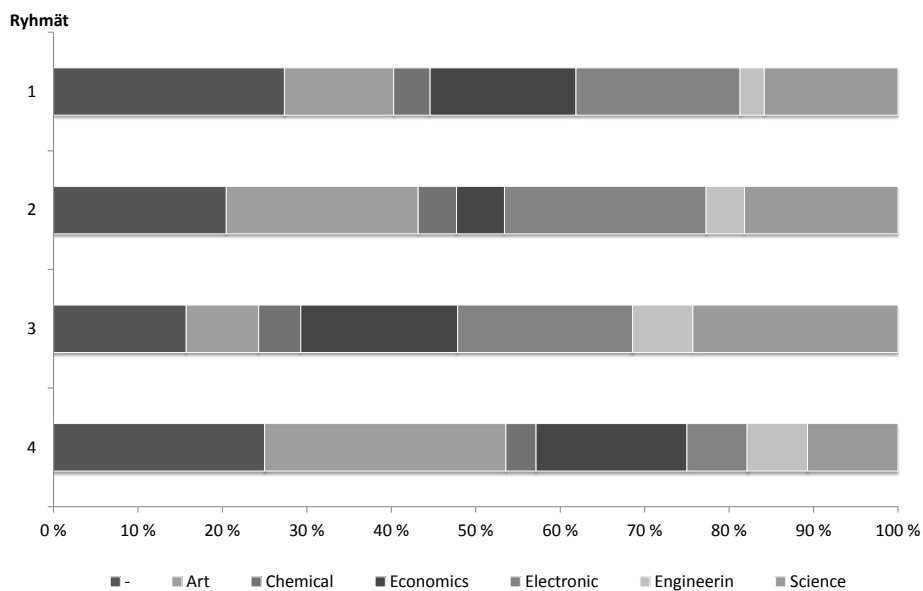
I do not 13 3.3 *

Male 255 64.9 *****

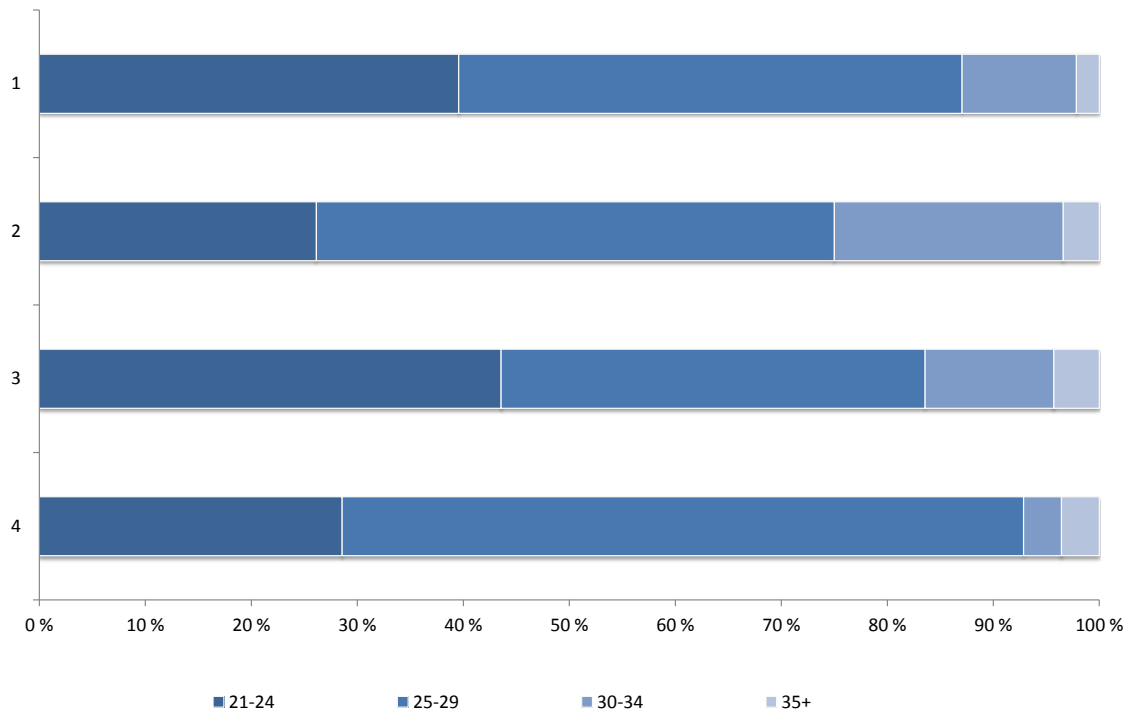
F Taustamuuttujia ryhmittäin



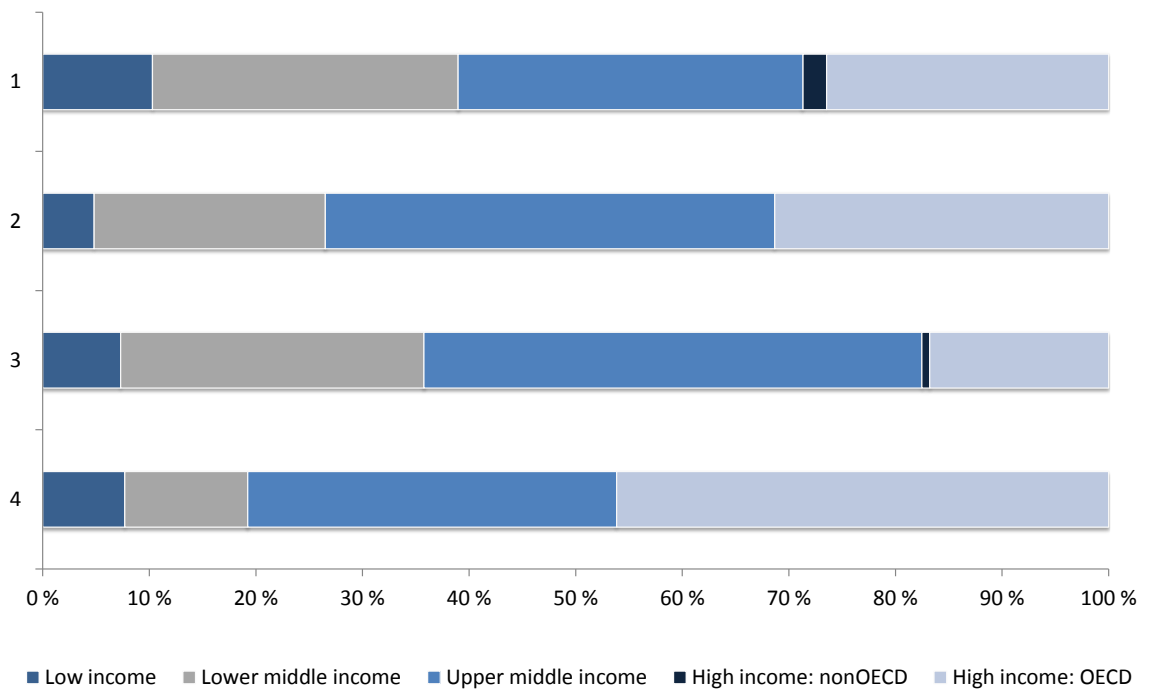
Kuvio F.1. Isän koulutusaste ryhmittäin



Kuvio F.2. Aalto-yliopiston koulut ryhmittäin



Kuvio F.3. Ikäluokat ryhmittäin



Kuvio F.4. Maailman pankin maiden tuloluokat ryhmittäin

G Ohjelmakoodit

Imputointi

imputointia.sps

```
DATASET ACTIVATE DataSet1.
DATASET DECLARE imputoitu.
MVA VARIABLES=V_1 V_2 V_3 V_4 V_5 V_6 V_7 V_8 FI_1 FI_2 FI_3 FI_4 FI_5 FI_6 FI_7
FI_8 FI_9 FI_10
FI_11 FI_12 FI_13 Aalto_1 Aalto_2 Aalto_3 Aalto_4 Aalto_5 Aalto_6 Aalto_7
Aalto_8 Aalto_9 Aalto_10
Aalto_11 Aalto_12 Aalto_13 VM_1 VM_2 VM_3 VM_4 VM_5 VM_6 VM_7 VM_8
Social_stand
/ID=ID
/REGRESSION(TOLERANCE=0.001 FLIMIT=4.0 ADDTYPE=RESIDUAL OUTFILE=imputoitu).
```

```
DATASET ACTIVATE imputoitu.
DESCRIPTIVES VARIABLES=V_1 V_2 V_3 V_4 V_5 V_6 V_7 V_8 FI_1 FI_2 FI_3 FI_4 FI_5
FI_6 FI_7 FI_8 FI_9
FI_10 FI_11 FI_12 FI_13 Aalto_1 Aalto_2 Aalto_3 Aalto_4 Aalto_5 Aalto_6
Aalto_7 Aalto_8 Aalto_9
Aalto_10 Aalto_11 Aalto_12 Aalto_13 VM_1 VM_2 VM_3 VM_4 VM_5 VM_6 VM_7 VM_8
/STATISTICS=MEAN SUM STDDEV VARIANCE MIN MAX SEMEAN.
```

```
DATASET ACTIVATE DataSet1.
DATASET DECLARE imputoitu2.
MVA VARIABLES=V_1 V_2 V_3 V_4 V_5 V_6 V_7 V_8 FI_1 FI_2 FI_3 FI_4 FI_5 FI_6 FI_7
FI_8 FI_9 FI_10
FI_11 FI_12 FI_13 Aalto_1 Aalto_2 Aalto_3 Aalto_4 Aalto_5 Aalto_6 Aalto_7
Aalto_8 Aalto_9 Aalto_10
Aalto_11 Aalto_12 Aalto_13 VM_1 VM_2 VM_3 VM_4 VM_5 VM_6 VM_7 VM_8
Nationality Dad_age Gender
/MAXCAT=25
/ID=ID
/CATEGORICAL=Nationality Dad_age Gender
/REGRESSION(TOLERANCE=0.001 FLIMIT=4.0 ADDTYPE=RESIDUAL OUTFILE=imputoitu2).
```

Faktorianalyysi

```
faktori tsii sti . txt
SURV0 98 edit field:      5001      10000      2000 (32 bit version)
00001 |*SAVE faktori t / mi ttausma lli , faktori anal yysi (data2)
00002 |*LOAD
00003 |*cd E: \e _jutut\anal yysi
00004 |*REDIM 10000,5000 /
00006 |*FILE SHOW DATAA
00007 |*FILE SAVE dataa .TXT TO NEW dataa / DELIMI TER=tab
00008 |*MASK=-----AA-AA-AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA-----

00011 |*file copy dataa to data2

00015 |*corr dataa
00016 |*facta corr .M, 5 CUR+2

00065 |*rotate FACT .M, 5 / ROTATI ON=VARI MAX

00070 |* /FTCOEFF AFACT .M, TFACT .M, MSN .M, FCOEFF .M
00071 |*Use FCOEFF .M for factor scores by LINCO <data>, FCOEFF .M(F1, F2, . . . )
00073 |*LINCO dataa, FCOEFF .M(F1, F2, F3, F4, F5)

00079 |* /loadfact dataa
```

Ryhmittely

```
klusterisiisti.sas
*****
Hierarki nen, kauimman naapurin menetelmä
*****
ods graphics on;
proc cluster data=isoE method=complete ccc pseudo print=15 outtree=puu6;
    VAR F3 F5 ;
    id id;
    * format country $12.;
run;
ods graphics off;

*4 klusteria?;

ODS PDF FILE =
'\ATKK\home\el\el kopone\Documents\Gradu\oi kei tatul oksi a\puu_6. pdf' ;
options vsize=9in hsize=6.4in htext=.9pct htitle=3pct;
axis1 order=(0 to 1 by 0.2);
proc tree data=puu6 out=uusi6 nclusters=4
    haxis=axis1 horizontal; *horizontal;
    title 'Hierarki nen ryhmi ttely, puukuvi o';
    height _rsq_;
    copy F3 F5 ;
    id id;
run;
ODS PDF CLOSE;

*****
Hierarki nen, lähimmän naapurin menetelmä
*****
ods graphics on;
proc cluster data=isoE method=single ccc pseudo print=15 outtree=puu6b;
    VAR F3 F5 ;
    id id;
    * format country $12.;
run;
ods graphics off;

*6 klusteria?;

ODS PDF FILE =
'\ATKK\home\el\el kopone\Documents\Gradu\oi kei tatul oksi a\puu_6b. pdf' ;
options vsize=9in hsize=6.4in htext=.9pct htitle=3pct;
axis1 order=(0 to 1 by 0.2);
proc tree data=puu6b out=uusi6b nclusters=6
    haxis=axis1 horizontal; *horizontal;
    title 'Hierarki nen ryhmi ttely, puukuvi o';
    height _rsq_;
    copy F3 F5 ;
    id id;
run;
ODS PDF CLOSE;

*****
k-means
*****
*FASTCLUS finds disjoint clusters of observations by using a k-means method
applied to coordinate data.;
proc fastclus data=isoIn out=Clust
    maxclusters=4 maxiter=100 LIST; *Output if the LIST parameter
is in the PROC FASTCLUS command, this table (only partially reproduced below)
shows the cluster assignment for each athlete (case).;
    var F5 F3;
run
```